THE PHYSICAL SOCIETY

REPORTS ON PROGRESS IN PHYSICS

VOLUME XII (1948–49)

Price to non-Fellows £2 2s. net (Postage 1s.)
Price to Fellows 25s. net (Postage 1s.)

Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Prince Consort Road
London S.W.7
1949

PRINTED IN GREAT BRITAIN

CONTENTS

	PAGE
Mass Spectrometry. By H. G. THODE and R. B. SHIELDS	1
Nuclear Paramagnetism. By B. V. ROLLIN	22
Phosphors and Phosphorescence. By G. F. J. GARLICK	34
Spontaneous Fluctuations. By D. K. C. MACDONALD	56
Recent Nuclear Experiments with High Voltage X-Rays. By W. BOSLEY and J. D. CRAGGS	82
Linear Accelerators. By D. W. FRY and W. WALKINSHAW	102
Viscosity and Related Properties in Glass. By G. O. JONES	133
Theory of the Oxidation of Metals. By N. CABRERA and N. F. MOTT	163
Fracture and Strength of Solids. By E. OROWAN	185
Multipole Radiation in Atomic Spectra. By A. RUBINOWICZ	233
Collisions between Atoms and Molecules at Ordinary Temperatures. By H. S. W. MASSEY	248
Low Temperature Physics. By K. MENDELSSOHN	270
Slow Neutron Absorption Cross Sections of the Elements. By M. ROSS and J. S. STORY	291
Molecular Distribution and Equation of State of Gases. By J. DE BOER	305
Author Index	375

AUTHOR INDEX

Abson, W., 76†
Adams, L. H., 157†
Adams, N. I., 104†
Akeley, E. S., 109†, 110
Albert, 221
Aldrich, L. T., 286†, 287†
Alekseyevsky, N., 274†, 277
Alfrey, T., 146, 150†, 152, 158
Alfrey, T., Jr., 230
Alkemade, C. Th. J., 77†
Allen, J. F., 280†
Allen, J. S., 5
Allen, W. D., 109†, 122†, 126†, 127†
Aller, L. H., 237†, 241†
Almy, G. M., 99†
Alpert, N. L., 28, 30†
Alvarez, L. W., 30†, 108, 128†
Amdur, I., 253, 254†
Anderson, E. E., 302 ref. (17)†, 303 ref. (67)†
Anderson, H. L., 30, 30† 302 ref. (12)†
Anderson, J. M., 260
Andrade, E. N. da C., 142, 196†, 197†
Andrew, E. R., 275
Andrews, D. H., 272†, 277, 277†
Andronikashvilli, E., 282
Araki, G., 233
Arnold, G. P., 303 refs. (55)† (66)†
Arnold, J. R., 302 ref. (43)†
Arnold, W. R., 30†
Aronin, L. R., 50†
Arrol, W. J., 16†, 18†
Aschermann, G., 274†
Aston, J. G., 273†
Aten, A. H. W., 86†
Atkins, K. R., 283
Atterling, H., 303 ref. (64)†

Bacher, R. F., 302 ref. (16)†
Bainbridge, K. T., 1†
Baker, C. P., 302 ref. (16)†
Baker, J. G., 237†, 241†
Baker, T. C., 199†
Baker, W. R., 130†
Bakker, C. J., 58†, 72, 86†, 235†
Baldwin, G. C., 84†, 85†, 86†, 89†, 90†, 91†, 92†, 93†, 95†, 99†
Balicki, S., 172†
Ballam, J., 130†
Ballini, R., 302 ref. (38)†
Bancroft, D., 147†
Band, W., 285†, 286, 287, 373†*
Barber, N. F., 2, 2†
Bardeen, J., 165†, 175
Barnes, R. B., 56†
Bartlett, J. G., 69†
Batchelor, R., 303 ref. (77)†
Bates, D. R., 176†, 262†
Bates, J. R., 258
Beams, J. W., 103†, 109†, 127†
Beck, P. A., 229†
Becker, 134
Becker, R., 228, 229, 265†
Becker, R. A., 89†
Beeman, W. W., 303 ref. (51)
Beers, Y., 235†
Bell, J., 71
Bénard, J., 171†, 172†, 178†
Bennett, R. D., 102†
Bennett, W. H., 7
Bennett, W. R., 64†
Bennewitz, K., 150†
Benton, A. F., 172†, 182†
Benzie, R. J., 33†*
Beringer, R., 108†, 109† 123†, 127†
Bernal, J. D., 135, 154†, 306
Bernstein, S., 302 ref. (20), 303 refs. (49)† (74)†
Berry, C. E., 14†
Berthelot, A., 302 ref. (38)†
Beth, E., 252†, 336†, 349†, 352†
Bethe, H., 236
Bethe, H. A., 236†
Betz, H., 180†
Beutler, H., 259
Bijl, A., 284†, 286† 343†
Birch, A. J., 273†
Birch, F., 147†
Birge, L., 346
Birkhoff, G. D., 318, 319, 319†
Birus, K., 48
Bishop, P. O., 76†
Bitter, F., 28†, 30†, 275†
Bittner, H. W., 272†
Blackman, M., 273†
Blaisse, B., 373†*
Blaisse, B. S., 343†
Blanchard, F. R., 272†
Blaton, J., 233, 233†, 235†, 240
Bleakney, W., 7†, 12†
Bleaney, B., 70†, 71†, 235
Blewitt, J. P., 104
Bloch, A., 150
Bloch, F., 22†, 23†, 25, 27†, 29†, 30†, 36, 334
Bloembergen, N., 23, 23†, 25, 27†, 28†, 29†, 31†, 65†, 77†
de Boer, J., 252†, 272†, 284†, 286†, 287, 337, 340, 342, 342†, 343†, 346†, 349†, 353†, 356†
Boggs, E. M., 373†*
Bogolubov, N., 284, 372*
Bonr, E , 303 ref. (64)†
Boltzmann, L., 308, 318
Booker, D. V., 302 ref. (7)
Boorse, J. A., 273†
Borchard, K. H., 146
Born, M., 58, 279†, 285†, 306†,, 307†, 314†, 316†, 318†, 323†, 331†, 337† 339†, 340†, 363†, 366
Borsellino, A., 284
Borst, L. B., 303 ref. (47)†
Bosley, W., 82, 85†, 97†, 98†
Botden, Th. P. J., 40†
Bothe, W., 82†, 89†, 302 refs. (21)† (37), 303 ref. (65)†
Bottema, M., 40†
Bouligand, G., 63
Bowen, E. G., 109†, 124†, 125†, 127†
Bowen, E. J., 53
Bowen, I. S., 233, 235†
Bowers, R., 283†
Boys, C. V., 147
Bradfield, G., 69†
Bradner, H., 128†, 130†
Brady, L., 245†
Bragg, W. L., 204, 204†
Brasefield, C. J., 37
Brattain, W. H., 165†
Brauer, P., 53
Bridgman, P. W., 201, 205, 206, 218
Brill, T., 303 ref. (60)†
Brillouin, L., 131
Brinkman, H. C., 240
Broadway, L. F., 255
Brobeck, W. H., 103†
Brochard, J., 241†
Brochard, P., 241†
Broer, L. J. F., 236, 236†, 243†, 278
Brown, H., 4†
Brown, J. B., 283†
Brown, L., 93†
Brown, R. C., 56
Bruck, G. G., 110†
Buckingham, R. A., 252†, 253†, 254, 254†, 256†, 352†, 353†
Buechner, W. W., 102†
Bünger, W., 35†, 43†
Bunting, E. N., 49
Burgers, J. M., 150
Burgess, R. E., 63, 70, 79
Burkhardt, G., 272
Burton, W. K., 173†, 179†
Busse, W. F., 230†
Butler, K. H., 51
Byk, A., 343

Cabrera, N., 172, 173†, 174†, 178, 179, 180†, 181†
Cairns, W. J., 203†, 209†, 210†, 214†, 216†, 217†
Caldirola, P., 284
Calvert, H. R., 263

† Joint author.
* Not referred to in text.

Campbell, N. R., 60, 61†, 70†, 73
Campbell, J. S., 235†
Campbell, W. E., 164†, 178†
Caratheodory, C., 318
Carroll, H., 302 r.f. (28)
Cas mir, H. B. G., 277†
Cavanagh, P. E., 302 ref. (7)†
Cazaud, R., 222
Chackett, K. F., 16†, 18†
Chadwick, J., 88†
Chandrasekhar, S., 58
Chapman, S., 250, 306–308†
Charlton, E. E., 82†, 103†
Cheng, K. C., 279†
Christensen, C. J., 71†, 76†
Chu, E. L., 112†
Clarke, C. L., 108†, 109†, 123†, 127†
Coates, W. M., 103†
Cockcroft, J. D., 102†
Coggeshall, N. D., 3, 14
Cohen, K., 8†
Collie, C. H., 89†
Collins, B. G., 88†
Collins, C. B., 10†, 15†, 16†, 303 ref. (78)†
Collins, S. C., 270
Collyer, P. W., 159
Colmer, F. C. W., 303 ref. (86)†
Coltman, J. W., 302 ref. (30)†
Condon, E. U., 7†, 240
Conklin, R., 93†
Considère, A., 188
Cook, D. B., 273†
Cook, W. R., 256†
Cooke, A. H., 33†*
Coon, J. H., 303 refs. (82)† (83)†
Coquelle, O., 171†
Cork, B., 130†
Corner, J., 306†
Corson, D. R , 84†
Coulson, C. A., 306†
Courant, E. D., 95
Cowhig, W. T., 113, 121†, 129
Cowling, T. G., 306–308†
Craggs, J. D., 97†, 98†
Crane, H. R., 96†, 97†
Crawford, R., 130†
Cullen, A. B., 109†, 127†
Curtis, B. R., 96†
Cuthbertson, C., 256†
Cuthbertson, M., 256†
Cutler, C. C., 112

Dain, J., 100†*
Dale, A. E., 144†
Dancoff, M., 233†, 243†
Darwin, C. G., 312†
Daunt, J. G., 63†, 79†, 273†, 278, 280†, 282†, 283†, 284†, 286†, 287
Davidenkov, N. N., 208
Debye, P., 368†, 372*
Dehlinger, U., 222
Delsasso, L. A., 97†
Demers, P., 302 ref. (33)†
Dempster, A. J., 1, 2, 303 refs. (52) (59) (61)†
Dent, B., 176†
Désirant, M., 273†, 275†, 278†
DeSorbo, W., 272† 277†
Destriau, G., 53
Deutsch, M., 245†
Deutschbein, O., 51, 236
Devons, S , 122
Devonshire, A F., 306†
Diaf, J. B., 303 ref. (74)†
Dingle, R. B., 288*
Dirac, P. A. M., 315
Diven, B. C., 89†
Dixit, K. R., 278
Doermann, F. W., 237, 237†
Dole, M., 8
Dolecek, R., 278†
Domb, C., 61, 68
Doran, R. F., 146†
Dorgelo, H. B., 260†
Dorrestein, S., 258†
Doty, P., 150†
Dresel, L. A. C., 72†
Druyvesteyn, M. J., 260†
Du Bridge, L. A., 4†
Duffendack, O. S., 258†, 259†
Dunning, J. R., 302 refs. (35)† (36)† (40)† (42)†
Dunworthy, J. V., 83
Durand, E., 233, 266, 267
Duykaerts, G., 281†
Dwyer, R. J., 266

Ebbinghaus, E., 260
Edlefson, N. E., 102†
Edlén, B., 234, 235
Egerton, A., 273†
Egglerc, 302 r.f. (13)†
Ehrenfest, P., 58, 309†, 312†, 318†, 321, 347†
Ehrenfest, T., 309†, 312†, 318†
Eichlin, C. G., 138†
Einstein, A., 56, 79, 349
Eisenschitz, R., 306
Eisenstein, A., 368†
Elder, F. R., 82†, 83†, 99†
Eley, D. D., 230†
Ellickson, R., 67
Ellickson, R. T., 52†
Elliot, L. G., 245†
Elliott, H. A., 193, 194
Elmer, R. A., 303 ref. (79)†
Endt, P. M., 77†
Enskog, D., 250
Eppstein, J. S., 303 ref. (77)†
Epstein, B., 197
Epstein, S., 16†, 18†
Estermann, I., 255†
Eucken, A., 265†
Evans, M. G., 258
Evans, U. R., 164, 172, 174, 228, 228†
Everhart, E., 109†, 122†
Ewald, W., 195†
Ewell, R. B., 254†, 306†, 343†, 346†
Ewing, D. H., 95†
Ewing, J. A., 222†
Eyring, H., 142, 263†, 306

Fairbank, H. A., 272, 282†, 286, 286†, 287, 287†
Fairbank, W. M., 282†
Falkoff, D. L., 244†, 245
Faraday Society, 35, 37
Feather, N., 68
Fenning, F. W., 302 ref. (8)† (22)†
Fermi, E., 302 refs. (12)† (14)†
Ferrien, J., 174†, 181†
Ferris, W. R., 72†, 73†
Fettweis, F., 207
Finch, G. I., 182†
Findlay, J. C., 282†
Fisher, J. C., 197†
Flechsig, W., 35†, 43†
Florenski, 198
Flowers, B. H., 303 r.f. (77)†
Fonda, G. R., 39, 49, 50†, 51, 52
Foner, S. N., 255†
Föppl, O., 221
Forrester, A. T., 5†
Foster, J., 71
Fowler, H., 303 ref. (81)†
Fowler, R. H., 154†, 306, 306†, 312†, 344
Fowler, W. A., 96†, 97†
Fox, R. E., 2†
Francis, V. J., 61†, 70†
Franck, J., 265†, 287
Frank, F. C., 163†, 173, 182†
Frank, J. V., 130†
Frank, N. H., 106†
Frankel, S., 110
Frankenhaeuser, B., 77†
Fränz, K., 65
Freed, S., 237, 237†
Frenkel, J., 23, 37, 135, 142, 156, 197†, 204
Frerichs, R., 235†
Fridman, J. B., 201, 218
Friedlander, G., 89†, 95†, 96†
Friedlander, H. N., 302 ref. (25)†
Friedman, H., 83†
Friis, H. T., 69
Froelich, H. C., 39, 49, 50†
Fröhlich, H., 153†
Frost, A. A., 266†
Fry, D. W., 82,† 104†, 109†, 118†, 124†, 125†, 127†
Fry, T. C., 61
Fuchs, K., 337†, 339†, 340†
Fünfner, E., 303 ref. (65)†
Furry, W. H., 67
Fürth, R., 56, 58, 65†, 74, 74†, 78, 306

Gabor, D., 108, 132*
Gaerttner, E. R., 97†
Gallop, J. W., 100†*
Garforth, F., 164, 174, 175

† Joint author.
* Not referred to in text.

Garlick, G. F. J., 35†, 37†, 38†, 40†, 41†, 43†, 44†, 45†, 46†, 47, 47†, 50†, 51, 51†, 52
Garrison, J. B., 275†
Geldermans, M., 343†
Gensamer, M., 218
Gentile, G., 284
Gentner, W., 82†, 89†
Gerjuoy, E., 237, 239
Ghering, J. G., 160†
Ghiorso, A., 303 ref. (73)†
Ghormely, J. A., 303 ref. (84)†
Gibbs, J. W., 320, 321, 365
Gibney, R. B., 274†
Gibson, A. F., 38†, 40†, 41†, 43†, 45†, 46†, 47†, 58†
Gibson, W. M., 88†
Gigli, A., 274†
Gilchrist, J., 223
Gillespie, A. B., 76
Gingrich, N. S., 368, 368†, 374†*
Ginlotto, J., 274†
Ginsburg, V., 278
Ginzberg, A., 109†
Ginzton, E. J., 106†, 107†, 108†, 109†, 110†, 114†, 118†, 120†, 121†, 123†, 125†, 126†, 127†
Gisolf, J. H., 36, 38†, 40†, 46
Glasstone, S., 263†
Gobrecht, H., 237
Goertzel, G., 244
Gogate, D. V., 286†, 288†*
Goldhaber, G., 88†
Goldhaber, M., 302 refs. (3)† (26)† (30)†, 303 refs. (62)† (85)†
Goldsmith, S. S., 303 ref. (79)†
Goldstein, L., 284, 303 ref. (89)†
Gooden, J. S., 103†, 109†, 124†, 125†, 127†
Gordon, H., 128†, 130†
Gordy, W., 235
Gorter, C. J., 22, 23, 23†, 31, 32, 236†
Gough, H. J., 222, 222†, 223†
Goudeket, M., 343†, 350†
Gow, J. D., 130†
Goward, F. K., 100†*
Graffunder, W., 260†
Graham, G. A. R., 301 ref. (1)†, 302 ref. (8)
Graham, R. L., 2†, 16†
Gran, W. H., 259†
Graves, A. C., 30†
Graves, J. D., 7†
Gray, L. H., 83
Grayson Smith, H., 282†
Green, H. S., 285†, 286, 306†, 307†, 314†, 316†, 318†, 323†, 328, 331†, 363†
Green, L. L., 88†
Green, T. D., 160†
Greiff, L., 8†
Greig, J. A., 109†, 127†
Greisen, K., 84†
Grenchik, R., 4†
Griffith, A. A., 192, 195, 198, 199
Grilly, E. R., 287†
de Groot, W., 36, 38†, 46, 260†, 343†
Gropper, L., 252, 332†, 333†, 352
Grosjean, C., 302 ref. (4)†
Grove, D. J., 4†, 12†, 13†
Grünewald, K., 171†, 172†
Guggenheim, E. A., 306†, 342
Gulbransen, E. A., 8†, 164†, 176†, 177†
Gunterschultze, A., 180†
Gurewitsch, A. M., 82†, 83†
Gurney, C., 155
Gurney, R. W., 36†, 40†, 46†, 135†, 168†, 171†, 181†
Guth, E., 88†
Gwathmey, A. T., 172†, 178†

de Haas, W. J., 253†, 272†
de Haas-Lorentz, G. L., 56
Haber, H., 268*
Hafstad, L. R., 86†
Hahn, W. C., 109†, 116†, 117†, 118†, 124†, 127†
Haigh, B. P., 231†*
Haines, G., 303 ref. (76)†
Halban, H., 89†
Hall, L. D., 303 ref. (79)†
Halpern, J., 97†, 109†, 122†, 275†
Halpern, O., 237†
Ham, W., 131†*
Hamilton, D. R., 244, 245
Hamilton, J., 252†, 353†
Hammel, E. F., 287†
Hamon, J., 174†, 180†, 181†
Hampton, W. M., 158
Hansen, W. W., 22†, 23†, 27†, 29†, 106†, 107†, 108†, 109†, 110†, 112†, 113†, 118†, 120†, 121†, 122, 124, 124†, 125†, 126†, 127†, 233
Hanson, A. O., 89†, 99†
Hanson, D., 222†, 223†, 229†
Hanstock, R. F., 222
Harkness, A. L., 2†
Harrington, E. L., 302 ref. (11)†
Harris, E. J., 76†
Harrison, S. F., 373†*
Hartman, P. L., 7†
Hartzler, A. J., 84†, 85†
R.-S. Harvie, R. B., 82†, 104† 106, 108, 109†, 110, 113, 115, 118†, 120, 124†, 125†, 126, 127†, 129
Hasbrouck, B., 303 ref. (47)†
Hassé, H. R., 256†
Hatton, J., 28†, 29†, 72†, 287†
Havens, W. W., Jr., 302 refs. (8)† (19)† (35)† (36)† (40)† (41)† (42)†, 303 refs. (45)† (46)† (69)†
Hawking, R. C., 19†
Haworth, L. J., 302 refs. (5)† (27)†
Haxby, R. O., 86†, 109†
Hayden, R. J., 19†, 20†, 303 refs. (57)† (63)† (68)† (88)†
Hayner, L. J., 75†
Hebb, M. H., 243†, 244†
Heisenberg, W., 278, 278†, 286, 286†
Heitler, W., 29†, 84, 88, 96, 97, 98, 99, 233
Hellingmann, J. E., 47†, 48†
Hellmann, H., 256
Hellwege, K. H., 236
Helmholz, A. C., 244
Hemmendinger, H., 52†
Hencky, H., 210
Hereford, F. L., 109, 124, 125, 127
Hereward, H. G., 302 ref. (7)† (10)†
Herman, R. C., 46†
Hess, D. C., 20†, 303 ref. (87)†
Hess, D. C., Jr., 303 refs. (57)† (63)† (68)†
Heymann, F. F., 85†
Heyn, F. A., 86†
Hickam, W. M., 4†, 12†, 13†
Hide, G. S., 103†
Hiedemann, E., 264
Hill, A. G., 50†, 235†
Hill, A. V., 77
Hill, C. G. A., 51
Hill, R., 210
Hilsch, R., 272
Hincks, E. P., 302 ref. (24)†
von Hippel, A., 40
Hipple, J. A., Jr., 2†, 4†, 7, 7†, 12†, 13†
Hirschfelder, J. O., 254†, 306†, 343†, 346
Hirzel, O., 94†
Hoagland, E. J., 302 ref. (44)†
Hodgins, J. W., 273
Holborn, L., 251†, 252†
Holland, A. J., 199†
Hollomon, J. H., 197†, 212†, 231*
Holtzman, R. H., 302 ref. (39)†
Holzworth, M. L., 229†
Hönig, R. E., 11
Hoogenstraaten, W., 40†
Hoogschagen, J., 236†
Hoover, J. I., 303 refs. (48)† (75)†
Hopkinson, R. G., 69
Horn, F. H., 277†
van Horn, J. R., 303 ref. (61)†
Horton, J. L., 7†
Hoselitz, K., 257
Houston, W. V., 241†
Hoyt, S. L., 214
Huber, O., 92†
Huddleston, C. M., 302 ref. (13)†

† Joint author
* Not referred to in text

Hudspeth, E. L., 109
Huff, L. D., 241†
Hughes, A. L., 2†
Hughes, D. J., 302 refs. (13)† (39)†
Hull, R. A., 273
Humblet, J., 233
Humfrey, J. C. W., 222†
Hundley, J. L., 15
Hunter, R. F., 19†
Hurst, D. G., 96†
Husimi, Kodi, 316, 332
Hutchison, C. A., 8†

Iitaka, I., 187, 187†, 230, 230†
Iliin, B., 251†, 252†
Inghram, M. G., 12, 14, 19†, 20†, 303 refs. (57)† (64)† (68)† (87)† (88)†
Inglis, C. E., 192, 202
Irvine, E., 142†
van Itterbeek, A., 251†, 267†, 271, 271†

Jackson, J. M., 265†
Jacobsohn, B. A., 25†, 30†
Jacquinot, P., 241†
James, E. G., 70†
Jeans, J. H., 308
Jenckel, E., 198
Jenkins, F. A., 237†, 239†, 240†
Jenks, G. H., 303 ref. (84)†
Jensen, P., 302 ref. (21)†
Joffé, A. F., 195, 195†
Johnson, A. F., 272
Johnson, J. B., 57, 76
Johnson, R. P., 36, 41
Johnston, H. L., 286†, 287†
Johnston, L. H., 132†*
Jominy, W. E., 229
Jones, E. J., 121†
Jones, G. O., 142, 145, 148, 158, 272†
Jones, H., 280, 280†, 283, 284
Jones, W. B., Jr., 303 refs. (54) (56)
Jordan, E. B., 1, 1†, 14†
Jordan, W. H., 303 ref. (75)†
Josephy, B., 259†
Justi, E., 274†

Kahn, B., 336, 337†, 340, 340†, 348, 349
Kamen, M., 9
Kantrowitz, A., 265, 266
Kapitza, P., 270, 280, 282
von Kármán, Th., 200
Karmarsch, K., 198†
Katcoff, S., 303 ref. (79)†
Keesom, P. H., 282†
Keesom, W. H., 251†, 252†, 280†, 281†, 282†, 353†, 354†, 368†, 369†
Kellogg, B. M., 235†
Kelvin, Lord, 318
Kennedy, W. R., 106†, 109†, 113†, 118†, 120†, 121†, 124–127†
Kerr, G. P., 258†, 263†
Kerst, D. W., 82, 83, 85, 99†, 103
Khintchine, A., 62
Kick, F., 200
King, L. D. P., 303 ref. (89)†
Kinman, T. H., 70†, 71†
Kirkwood, J. G., 306, 308, 321–323, 332, 334, 349, 358, 363, 366, 373†*
Kirpitschewa, M. W., 195†
Kisilbasch, B., 258†
Kistemaker, J., 283, 353†
Klaiber, G. S., 84†, 85†, 86†, 90–93†, 95†, 99†
Klasens, H. A., 41, 41†, 45†
Kleen, W., 72
Klemens, P. G., 68
Kneser, H. O., 264
Koch, H. W., 85†, 89†
Kommers, J. B., 221†, 222†
Kompfner, R., 63†, 71, 72†, 74†, 79†
Kondratjew, V., 258†
König, H., 144, 147
Kontorova, T. A., 197, 197†
Koopman, B. O., 319
Koppe, H., 279
Kraak, H. H., 251†, 252†, 353†, 354†
Kramers, H. A., 233, 233†, 234, 234†
Kramers, H. J., 373*
Kröger, F. A., 37–40, 40†, 47, 47†, 48†, 49, 51, 51†
Kröger, J. H., 37, 37†, 38, 38†, 40, 40†
Kronig, R. de L., 23†
Kruger, P. G., 97†
Kruithof, A. A., 260†
Krupkowski, A., 172†
Küchler, L., 265
Kuhn, W., 150
Kuntze, W., 214
Kurrelmeyer, B., 75†
Kurti, N., 31†
Kvifte, H., 260†

Laidler, K. J., 263†
Laidler, K. M., 264
Lamont, H. R. L., 110†
Landau, L., 262, 264†, 265†, 284, 287†
Landgraf, F. K., Jr., 230†
Lane, C. T., 282†, 286†, 287†
Langevin, M. P., 58, 373*
Langmuir, R. V., 82†, 83†, 99†
Lantz, P. M., 303 ref. (57)†
Lanzl, L. H., 99†
Lapointe, C., 302 ref. (23)†
Lapp, R. E., 303 ref. (61)†
Larrick, L., 230†
Larsen, A. H., 272†
Lasarew, G. B., 22†
Lasich, W. B., 97†
Laslett, L. J., 96†
von Laue, M., 237, 278†
Laughlin, J. S., 99†
Laumann, E., 275†
de Launay, J., 278†
Laurence, G. C., 83, 302 ref. (10)†
Lauritsen, C. C., 97†, 102†
Lavatelli, L. S., 302 ref. (17), 303 ref. (67)†
Lawrence, E. O., 102†, 103†, 130†*, 132†*
Lawton, E. J., 109†, 116†, 117†, 118†, 120, 122, 124†, 127†
Lax, M., 84, 88
Leary, J. A., 303 ref. (79)†
Lees, W. S., 259†
Lehr, C. G., 28†, 30†
Leighton, P. A., 263†
Leipunsky, A., 258†
Lennard-Jones, J. E., 176†, 256, 306†, 344, 345
Lessig, E. T., 230†
Levi-renz, H. W., 47, 49, 50, 51
Levinthal, E. C., 30†
Levitsky, M. A., 195†
Lewis, L. G., 6†, 20†
Lewis, W. B., 79, 84
Lichtenberger, H. V., 303 ref. (60)† (81)†
Lienhard, O., 92†
Lifschitz, E., 285
Lillie, H. R., 138, 142, 143, 157
Lin, S. T., 28†, 30†
Lindh, A. E., 243
Ling, D. S., 244†
Linwood, S. H., 49†
Lipson, H. C., 258†, 263†
Littleton, J. T., 138†, 157
Livesey, D. L., 88†
Livingstone, M. S., 130†
Loach, B., 112†, 118†, 121†
Lockwood, J. A., 108†, 109†, 123†, 127†
Lomer, W. M., 204†
London, F., 262, 279, 283, 284, 286, 287†, 349
London, H., 274, 286
Lorentz, H. A., 321
Lossing, F. P., 10†, 12†, 13†, 15†, 303 ref. (78)†
Loughborough, D. L., 230†
Lozier, W. W., 12†
Lubahn, J. D., 215†
Ludwik, P., 204, 206, 214†
Luebke, E. A., 95† 302 refs. (5)† (27)†
Lunbeck, R. J., 343†

McAdam, D. J., Jr., 206†, 215†, 228
McCarthy, R. L., 108†, 109†, 123†, 127†
McDaniel, B. D., 302 refs. (16)† (17)†, 303 refs. (53) (67)†

† Joint author.
* Not referred to in text.

MacDonald, D. K. C., 56, 65, 65†, 66, 70, 74, 74†, 77†, 79, 272†, 273†, 277†
Macfarlane, G. G., 76
Mack, C., 158
McKeag, A. H., 50†
McKenzie, K. R., 103†
McMillan, E. M., 103, 103†
McNamara, E. P., 146†, 156†
Macnamara, J., 10†, 15†, 16†, 303 ref. (78)†
Magat, M., 266†
Magee, J. L., 263†
Magnusson, L. B., 303 ref. (73)†
Maier, A. F., 218
Manjoine, M. J., 190†
Manley, J. H., 302 ref. (5)† (27)†
Mann, W. B., 19†
Margenau, H., 253
Marhoefer, C. J., 88†
Mariens, P., 267†, 271†
Marshall, J., 302 refs. (14)† (15)
Marshall, L., 302 ref. (14)†
Marshall, L. C., 128†
Martindale, J. G., 196†
Martinelli, F. A., 132†*
Massey, H. S. W., 176†, 249†, 250†, 251†, 252†, 253†, 256†, 257†, 262†, 352† 353†
Mattauch, J., 1
Maurer, W., 259†, 303 ref. (70)
Maxwell, E., 275†
Maxwell, J. C., 161*, 318
May, A. N., 302 ref. (24)†
Mayer, J. E., 306†, 339†, 340†, 346†, 356†, 373, 373†
Mayer, M. G., 306†, 339†, 340†
Meaker, C. L., 000†
Mebs, R. W., 206†, 215†
Meijer, R. R., 303 ref. (90)
Meissner, K. W., 260†
Meitner, L., 303 ref. (71)
Melkonian, E., 302 ref. (36)†
Mellink, J. H., 281, 281†, 283†
Mendelssohn, K., 272†, 273, 273†, 274, 275, 277†, 278, 280†, 281, 282†, 283, 283†, 284, 284†, 286, 286†
Menke, H., 368†
Menzel, D. H., 237†, 241†
Merrill, J. B., 51†
van der Merwe, J., 163, 163†, 182, 182†
Merwin, H. E., 138†
Meshkovsky, A., 277†
Metzger, F., 245†
Meyer, H. F., 46†
Meyer, L., 280†, 281†, 283†, 285†
Meyer, R. C., 86†
Meyerott, R., 257
Michels, A., 252†, 284†, 286†, 342†, 343†, 346†, 349†, 350†, 353†, 356†, 373†*
Michels, C., 343†, 373†*
Michelson, A. A., 145
Migunov, L., 274†
Milatz, J. M. W., 63†, 69†, 77†, 79†
Milianczuk, B., 237
Millman, S., 235†
Mills, B. Y., 109, 125, 126, 127
Milton, R. M., 272†, 277†
Mitchell, A. C. G., 258†, 259†, 260†
Moak, C. D., 303 ref. (75)†
Moffatt, J., 72
Mohr, C. B. O., 250†, 251†, 256†, 257†
Monk, G. W., 7†
Montandon, R., 217†
Montgomery, C. G., 108†, 109†, 123†, 127†
Montroll, E. W., 340†, 346†, 356†
Moore, H. F., 221†, 222†
Morey, G. W., 138†, 139
Morrison, P., 233†, 243†
Morrison, P. J., 103
Mott, N. F., 36†, 40†, 46†, 135†, 163, 166, 167, 168, 168†, 171†, 175, 176, 178, 180, 181†, 183, 249, 249†, 262†, 265†
Moullin, E. B., 76
Moyer, B. J., 303 ref. (52)†
Moyer, W. E., 302 ref. (34)†
Mrozowski, S., 233, 237†, 238, 239, 239†, 240, 240†
Muehlhause, C. O., 83, 302 ref. (26)†
Mullett, L. B., 82†, 104†, 108, 109, 112†, 118†, 121†, 124†, 125†, 127†
Munn, A. M., 302 ref. (10)†
Murgatroyd, J. B., 146†, 147, 147†, 160
Murphy, B. F., 8†
Myers, F. E., 88†

Nadai, A., 190†, 218
Nagle, D. E., 30†
Nail, N. R., 52†
Nash, W. F., 97†, 98†
Nelson, E., 243†, 244†
von Neumann, J., 315, 319
Newbery, G. R., 109†, 113, 117†, 118, 121†, 122†, 127†
Newbury, E. G., 303 ref. (79)†
Newton, R. F., 138†
Nicodemus, D., 30†
Nicoll, F. H., 51
Nielsen, C. E., 4
Niemi, N. J., 302 ref. (7)†
Nier, A. O., 2, 6, 8†, 12, 286†, 287†
Niessen, K. F., 58†
Niewodniczański, H., 233, 235, 235†
Nijhoff, G. P., 251†, 252†
Nobles, R. A., 303 ref. (82)†
Norris, L. D., 19†
Norrish, R. W., 259†, 263†
North, D. O., 69, 72†, 73†
Norton, K. A., 69†
Novick, A., 30†
Noyes, B., Jr., 146†
Noyes, W. A., 263†
Nutting, J., 213, 230
Nye, J. F., 203†, 209†, 210† 214†, 216†, 217†
Nyquist, H., 66

Oblad, A. G., 138†
Obreimoff, J. W., 199
Ogg, R. A., 273, 277†
Ogle, W., 93†, 97†
Oldenberg, O., 266†, 267,. 267†
Olesen, R. C., 13
Oliner, A. A., 110
Olink, J. Th., 77†
Oliphant, M. L., 103†
Olsen, L. O., 258, 258†, 259,. 263†
Omberg, A. C., 69†
O'Neal, R. D., 303 ref. (62)†
Onsager, L., 358
Oppenheimer, F., 128†, 130†,. 132†
Oppenheimer, J., 96†, 97†
Oppenheimer, J. R., 347†
Orlin, J. J., 99†
Ornstein, L. S., 60, 63†, 69†, 312, 364†, 365†, 367†
Orowan, E., 188, 190–196, 198–202, 203†, 204, 209,. 209†, 210, 210†, 214†, 215†, 216†, 217†, 221, 223,. 227–229
Osborne, C. L., 303 ref. (47)†
Osborne, D. V., 282
Otto, J., 251†, 252†
Owens, J. S., 258†

Packard, M. E., 22, 23, 27, 29,. 30, 30†
van Paemel, O., 271†
Page, L., 104†
Pake, G. E., 31
Palevsky, H., 4†
Paneth, A. R., 302 ref. (10)†
Panofsky, W. K. H., 128,. 130†, 132*
Pardue, L., 303 ref. (75)†
Parker, E. R., 229
Parker, G. W., 303 ref. (57)†
Parker, W. L., 52†
Parks, G. S., 142†
Pasternack, S., 241
Pathak, P. D., 289†*
Pauli, W., 313
Pawsey, J. L., 65, 65†
Payne, R. M., 98†
Peacock, W. C., 245
Pearlman, D., 52†
Pearlman, H., 254†
Pearson, G. L., 71†, 76†,. 274†
Pearson, K., 68
Peierls, R., 35
Peirce, F. T., 197
Pellam, J. R., 282, 282†

† Joint author.
* Not referred to in text.

Penning, F. M., 260†
Pepper, D. C., 230†
Perlman, M. L., 89†, 95†, 96†
Peshkov, V., 282, 285
Petch, N. J., 202, 212
Peters, B., 303 ref. (52)†
Peterson, K. C., 303 ref. (49)†
Philbert, G., 278
Phillips, P., 146
Pickard, G. N., 271†
Pierce, J. R., 71, 75, 75†
Pippard, A. B., 274
Pirenne, M. H., 74, 79
Pitt, A., 282†
Pitzer, K. S., 342
Plancherel, M., 318†
Planck, M., 79
Plutonium Project Report, 15
Pluvinage, P., 233, 241, 241†
Poincaré, H., 318
Polanyi, M., 192, 195†
Pollock, H. C., 82†, 83†, 99†
Pomerance, H., 303 ref. (75)†
Pomerance, H. S., 303 ref. (91)
Pomerance, M., 303 ref. (48)†
Pomeranchuk, I., 287†
Pontius, R. B., 273†, 277†
Poss, H. L., 28†, 30, 30†
Post, R. F., 103
Pound, R. V., 22†, 23†, 26†, 27–31, 27†, 28†, 29†, 31†, 65†
Povelties, J. J., 303 ref. (79)†
Powell, C. F., 257†
Prandtl, L., 210
Preston, F. W., 199†
Prileschajewa, N., 258, 258†
Prins, J. A., 306, 364†, 367†
Probst, R. E., 286†, 287†
Pulley, O. O., 109†, 124†, 125†, 127†
Purcell, E. M., 22†, 23†, 26†, 27†, 28†, 29†, 31†, 65†

Quarrell, A. G., 182†

Rabi, I. I., 255, 303 ref. (69)†
Rademakers, A., 277†
Ragazzini, J. R., 64
Rai, R. N., 286†
Rainwater, J., 302 refs. (8)† (19)† (35)† (36)† (40)† (41)† (42)†, 303 refs. (45)† (46)† (69)†
Rall, W., 1†, 20
Ranby, P. W., 50†
Randall, J. T., 35, 35†, 40, 41–43†, 47†, 48†, 50†, 51
Rankama, K., 9
Ranque, G., 272
Rapuano, R. A., 109†, 122†
Rasetti, F., 302 refs. (23)† (31)
Ratcliffe, J. A., 65, 65†
Rayleigh, Lord, 68
Reel, H. L., 302 ref. (4)†
Reimann, A. L., 36, 46
Reinkober, O., 197, 198
Reynolds, J. H., 303 ref. (88)†
Reynolds, J. M., 274†
Rice, O. K., 287†
Rice, P. J., 108†, 109†, 123†, 127†
Rice, S. O., 63, 64, 64†, 67
Richards, W. T., 264, 265
Richman, C., 128†, 130†, 132†
Richtmyer, R. D., 110†
Riddiford, L., 97†
Riehl, N., 35, 36, 40
Rieke, F. F., 267†
Rivlin, R. S., 61
Roberts, A., 27, 30, 30†, 235†, 245†
Roberts, R. B., 86†
Roberts, W. L., 76†
Robinson, C. S., 85†
Robinson, D. M., 148
Robinson, H. A., 241, 241†
Rock, S. M., 12†, 13†
Roebuck, J. R., 254†, 306†, 343†, 346†
Rogers, D. C., 72†
Rollin, B. V., 26, 28†, 29†, 31, 33*, 287†
Rosa, A. M., 236
Rose, M. E., 267†
Rosenthal, A., 318†
Roseveare, W. E., 343†, 346†
Rosin, S., 255†
Rossi, B., 84†
Rössler, F., 266, 267
Rötger, H., 147, 150†, 153
Rothschild, S., 39, 47, 49
Rowland, E. N., 61
Roy, A. S., 267†
Ruark, A. E., 213†
Rubinowicz, A., 233†, 240
Rupp, H., 48
Rushbrooke, C. S., 306†
Russell, S. B., 207
Ryde, J. W., 70†, 71†

Sachs, G., 215†
Sachs, R. G., 30, 30†
Sack, R., 153†
Sack, R. A., 193
Sakata, S., 83†
Sakharov, P. S., 212
Salvetti, C., 284
Sargent, B. W., 302 refs. (7)† (10)†
Saris, B. F., 280†
Sawyer, R. B., 303 ref. (49)†
Schäfer, K. L., 350
Schafroth, R., 243
Schallamach, A., 272
Scheffers, H., 256†
Scherrer, P., 92†
Scheu, R., 214†
Schiff, L. I., 103, 104, 237, 284
Schmid, E., 195†
Schmidt, F. H., 303 ref. (52)†
Schneider, 72†
Schön, M., 40, 41
Schottky, W., 73, 76
Schubert, G., 284
Schubnikow, L. W., 22†
Schulman, H. J., 51†
Schulman, J. H., 38, 49
Schultz, H. L., 108†, 109†, 123†, 127†
Schulz, L. G., 302 ref. (3)†
Schurkow, S., 198
Schwarz, M., 41†
v. Schweidler, E., 73
Schwinger, J., 30†
Seaborg, G. T., 303 ref. (72)
Segrè, E., 233, 235†, 302, ref. (6)
Seitz, F., 35, 40, 43, 46, 47
Selényi, P., 236, 237
Seligman, H., 302 refs. (8)† (22)†
Serber, R., 103†, 128†, 130
Seren, L., 302 refs. (25)† (34)†
Sewell, D. C., 103†
Shalnikov, A. I., 274†, 277†
Shapiro, M. M., 303 ref. (74)†
Sharrah, P. C., 369†
Sharvin, J. V., 274†
Shaw, A. E., 1†, 303 ref. (63)†
Shearin, R. E., 213†
Sherman, J., 146†, 156†
Shields, R. B., 12†, 13†
Shockley, W., 75†, 165†
Shoenberg, D., 274, 275†, 278†
Shortley, G. H., 237†, 241, 241†
Shoupp, W. E., 86†
Siebel, E., 206
Siegbahn, M., 243
Siegfried, W., 229
Sienko, M. J., 277†
Sigurgeirsson, Th., 303 ref. (64)†
Silverman, S., 56†
Simha, R., 150
Simon, F., 31, 31†, 135, 137, 271, 272
Simon, F. E., 271†, 272†
Simpson, K. M., 103†
Siri, W., 5
Skaggs, L. S., 99†
Skinner, H. W., 259†
Slack, M., 68
Slater, J. C., 106†, 109†, 106–110, 112, 113, 116, 117, 120, 122, 122†, 123, 127, 251, 275, 275†, 330, 332, 353
Sloan, D. H., 103†, 132*
de Smedt, J., 368†, 369†
Smeets, C., 302 ref. (38)†
Smekal, A., 139, 152, 195
Smid, E., 195†
Smit, J. H., 258†
Smith, C. L., 153
Smith, J., 319†
Smith, L. G., 12†
Smith, L. P., 7†
Smith, P. T., 12†
Smith, R. A., 76, 257†
Smith, R. N., 109†
Smith, S. R., 8†
Smith, W. M., 259†, 263†
v. Smoluchowski, M., 58, 374*

† Joint author.
* Not referred to in text.

Smyth, C. N., 72
Snoddy, L. B., 103†, 109†, 127†, 131†*
Sohncke, L. 191
Sommerfeld, A., 284
Spaght, M. E., 142†
Spedding, F. H., 236†
Spence, R. W., 30†
Spenke, S., 73†
Sperry, P. R., 229†
Spiers, J. A., 32
Squire, C. F., 275†, 282†
Stanford, C. P., 303 ref. (74)†
Stanley-Smith, G., 218
Stanworth, J. E., 143, 144†, 154
Stark, J., 256†
Starr, A. T., 109
Staub, H. H., 30†
Steinheil, A., 174
Stephens, W. E., 2†, 7, 86†
Stephens, W. H., 19†
Stephenson, T. E., 303 ref. (74)†
Stern, O., 255†
Stevenson, A. F., 240
Stevenson, D., 306†
Stever, H. G., 83
Stewart, D. W., 8†
Stewart, J. L., 302 ref. (11)†
Stott, V. H., 138, 142†, 145
Stout, J. W., 278
Stratton, J. A., 107
Straus, H. A., 11
Strong, J., 68
Strong, J. D., 303 ref. (75)†
Strutt, M. J. O., 69†, 70†, 71†, 72†
Stueckelberg, E. C. G., 262
Sturm, W., 302 ref. (34)†
Sturm, W. J., 303 refs. (55)† (58) (66)†
Sugarman, N., 302 refs. (43)† (44)†
Sunyar, A. W., 303 ref. (85)†
Susken, I., 258†*
Sutton, R. B., 302 ref. (17)†, 303 ref. (67)†
Swank, R. K., 4†
Sweeton, F. H., 303 ref. (84)†
Swenson, C. A., 272†
Swift, H. W., 219
Swings, P., 235, 235†
Sydoriak, Sg., 287†
Sykes, R. F. R., 146†, 147†
Symonds, J. L., 109†, 122†, 126†, 127†
Szasz, G. J., 273†

Taconis, K. W., 365†
Talbot, J., 172†, 178†
Tammann, G., 135, 135†
Taylor, G. I., 182
Taylor, N. W., 146, 146†, 156†
Tchorabdji-Simnad, M., 228†
Teller, E., 29†, 264†, 265†
Terenin, A., 258†
Terrien, J., 174†, 181†
Thode, H. G., 2†, 8†, 10†, 12†, 13†, 15†, 16†, 303 ref. (78)†
Thomas, G., 303 ref. (80)†
Thomas, U. B., 164†, 178†
Thornton, R. L., 103†
Thosar, B. V., 51
Tiede, R. L., 153†
Tisza, L., 280, 284, 287
Tolman, R., 306, 309, 365
Tool, A. Q., 138, 138†, 155, 157, 159
Tooley, F. V., 153†
Torrey, H. C., 22†, 23†, 26†
Toulis, W., 132†*
Triede, R. L., 153†
Trimble, R. M., 213†
Trotter, H., Jr., 103†
Trump, J. G., 102†
Truter, E. V., 273†
Tsien, L. C., 196†, 197†
Tumanov, K. A., 274†
Turin, J. J., 96†
Turkel, S. H., 302 ref. (25)†
Turner, C., 128†, 130†
Turner, D., 142†
Turner, W. E. S., 199†
Twyman, F., 157
Tyndall, A. M., 256, 257†

Ubbink, J. B., 253†
Uhlenbeck, G. E., 63†, 243†, 244†, 252†, 313, 332, 332†, 333, 333†, 336†, 337†, 340†, 349, 349†, 352†
Ullrich, E. H., 72†
Ulrich, A. J., 303 ref. (47)†
Urbach, A., 41†
Urbach, F., 41†, 52†, 53
Urey, H. C., 8, 8†, 10, 10†
Ursell, H. D., 336

Valasek, J., 138†
Van Atta, L. C., 88†
Van de Graaff, R. J., 102, 102†
Vanderberg, R. M., 11
Van Vleck, J. H., 236
Vaupel, O., 195†
Vegard, L., 260†
Veksler, V. K., 103
van der Velden, H. A., 63†, 79†
Versnel, A., 72†
Verwey, E. J. W., 180
Villars, F., 30
Voigt, W., 150
Volz, H., 68, 302 ref. (32)†
de Vries, Hl., 79

Waddell, J. S., 303 ref. (79)†
Wäffler, H., 92†, 94†
Wagner, C., 171†, 172†
Walcher, W., 6
Waldman, B., 88†
Walker, D., 246†
Walker, R. L., 301 ref. (2)†
Walkinshaw, W., 82†, 106, 107, 110, 112, 118
Wall, C. N., 306
Wallace, J. R., 302 ref. (39)†
Waller, I., 29
Wallman, M. H., 265
Walsh, K. A., 303 ref. (79)†
Walstra, W. K., 251, 252, 353†, 354†
Walton, E. T. S.; 102†
Wang, C. C., 129
Wang, M. C., 63†
Wangsness, R. K., 25†, 30†
Warburg, E., 145
Ward, A. G., 68, 142, 154
Ward, A. H., 246†
Ward, R., 47, 52
Warren, B. E., 134
Washburn, H. W., 12†, 13†, 14†
Washington, H. S., 260†
Watson, D. S., 69†
Watson, W. W., 108†, 109†, 123†, 127†
Wattenberg, A., 302 ref. (12)†, 303 refs. (80)†, (81)†
Way, K., 000†
Webber, R. T., 274†
Weibull, W., 197
Weidmann, G., 145, 148
Weil, G. L., 302 ref. (12)†
Weisskopf, V. F., 95†
Weissman, S. J., 237†
Weisz, P. B., 7
Welch, H. W., 109†
Wells, A. F., 47, 48
Wells, W. H., 86†
Wergeland, H., 250
Westendorp, W. F., 82†, 103†
Westermijze, W. K., 280†
Wetmore, W. L., 138†
Wexler, A., 277
Weyl, W. A., 49†, 162*
Whaley, R. M., 109†
Whalley, W. B., 5†
Wheeler, M. A., 229†
Whitehead, S., 147
Whitehouse, W. J., 301 ref. (1)†
Whittaker, A., 303 ref. (77)†
Wiblin, E. R., 303 ref. (86)†
Wicker, E. R., 110†
Wideroe, R., 103
Wiechert, E., 149
Wiedenbeck, M. L., 88, 88†
Wiener, N., 62
Wigner, E., 259, 332
Wiley, H. F., 12†, 13†
Wilhelm, J. O., 282†
Wilkins, M. H. F., 35†, 37†, 40†, 41†, 42†, 43†, 44†, 45†, 46†, 47†, 48†, 50†, 51†
Williams, T. W., 2†
Williamson, E. D., 157†
Willshaw, W. E., 109†, 110†, 113, 116, 117†, 118, 121, 122†, 127†
Wilson, C. W., 83
Wilson, R., 65, 89†
Wilson, R. R., 84†
Winans, J. G., 258

† Joint author.
* Not referred to in text.

Winter, A., 139, 159
Wise, M. E., 41†, 45†
Wohl, K., 266†
Wöhler, A., 221
Wolf, R., 259†
Wollan, E. O., 303 refs. (49)† (75)†
Woodward, P. M., 61, 67
Woodyard, J. R., 103, 104, 128†, 130†, 132†*
Worden, C. O., Jr., 230†
Wouters, H., 343†
Wu, C. S., 302 refs. (35)† (40)† (41)† (42)†, 303 ref. (45)†

Wysong, W. S., 164†, 176†, 177†

Yamagishi, I., 187†, 230†
Yang, C. N., 246
Young, T., 192
Yukawa, H., 83†
Yvon, J., 233, 241, 285, 314, 325, 340, 364, 366–368

Zachariasen, W. H., 134
Zapffe, C. A., 230, 230†
Zavoisky, E., 23
Zemansky, M. W., 258, 258†, 259†, 260, 260†, 263, 273†
Zener, C., 140, 212†, 262, 264, 265
Zernike, F., 364†, 365†, 367†
Ziegler, M., 75
Ziegler, W. T., 277†
v. d. Ziel, A., 69, 70, 71, 72, 72†, 78
Zinn, W. H., 302 ref. (12)†, 303 ref. (50)
v. Zolingen, J. J., 78
Zschokke, H., 217†
Zwicky, F., 192

MASS SPECTROMETRY

By H. G. THODE and R. B. SHIELDS
McMaster University, Hamilton, Ontario

CONTENTS

	PAGE
§ 1. Introduction	1
§ 2. General instrumentation	2
2.1. Direction focusing	2
2.2. A.C. operated instruments	3
2.3. Ion detecting and measuring methods	4
2.4. Ion sources	5
2.5. New mass spectrometers	7
§ 3. Relative abundances of isotopes	8
3.1. Thermodynamic properties of isotopic substances	8
3.2. Variations in isotopic content of boron and sulphur	10
3.3. High precision instruments	10
§ 4. Absolute abundance measurements	13
§ 5. Nuclear problems	15
5.1. Fission yields by the mass spectrometer method	17
5.2. Half life measurements	19
5.3. Assignment of mass to fission isotopes	19
References	20

§ 1. INTRODUCTION

Mass spectrographs were first designed to identify isotopes and measure their abundances. In recent years special instruments have been designed for precision mass determinations (Dempster 1935, Bainbridge and Jordan 1936, Mattauch 1936, Jordan 1940, Shaw and Rall 1947). To determine chemical atomic weights, and in order to make certain nuclear calculations, both the abundances of the isotopes and the isotopic masses must be known. However, because of the widely different applications and different types of instruments used, it has been convenient to divide the field into two parts: mass spectroscopy, which has to do with photographic recording instruments for precision mass determinations, and mass spectrometry, which has to do with instruments for ion collection and measurement in connection with relative abundance determinations. It is the purpose of this report to review recent developments in the latter part of the field.

The mass spectrometer is no longer a research instrument for the pure physicist alone. Ever since it became possible to separate isotopes, mass spectrometers have been needed for routine work in physics, chemistry, and biology. The number of applications of mass spectrometry are ever-increasing and to-day several commercial concerns are manufacturing instruments to meet the growing demand. The mass spectrometer has been used successfully to study the stable isotopes of several elements produced in the fission of uranium. These isotopes, which are end products of fission decay chains, are formed according to the fission yield of the respective mass chains. The mass spectrometer permits the identification of stable and long lived isotopes produced in fission, the determination of mass numbers of fission chains and the half lives of fission products, and finally relative fission yields. A great amount of work remains to be done in this field.

Instruments of high precision have opened up another field of investigation. Due to differences in thermodynamic properties of isotopic substances, small variations in isotopic abundances occur in nature. These small variations in isotopic ratios which are of interest to physicists, chemists, geologists etc. can now be measured with considerable accuracy. Mass spectrometers are also used in the field of gas analysis. Each compound introduced into a gas type mass spectrometer yields a mass pattern depending on the mass fragments formed under the conditions of operation. A study of these patterns gives fundamental information about molecular systems and makes possible a gas analysis scheme.

Considerable progress has been made in these various fields of mass spectrometry in the past few years. Instruments and techniques have been developed to handle specific problems. In this report these new techniques will be discussed as well as some of the more important results obtained.

§2. GENERAL INSTRUMENTATION

2.1. *Direction Focusing*

Although there are a number of arrangements which possess mass selective properties and which might conceivably form the basis for mass spectrometers, the use of direction focusing instruments has become universal for the measurement of relative abundances (Fox, Hipple and Williams 1944, Nier 1947, Graham, Harkness and Thode 1947). In general, ions which are produced at the source, usually by electron impact, form a beam upon being accelerated by a high voltage V, after which they are deflected by a uniform magnetic field H, according to the relation $e/m = 2V/H^2R^2$, where e is the ionic charge, m is the mass, and R the radius of curvature. By a suitable adjustment of the parameters V or H, ions of

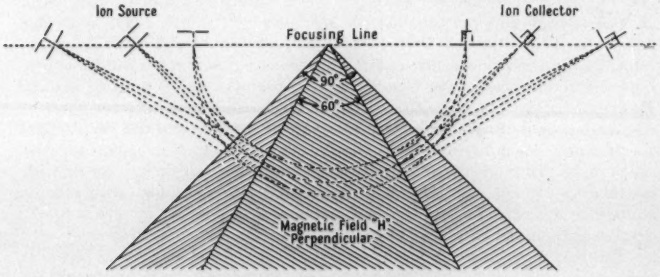


Figure 1. Direction focusing.

any desired mass may be collected by a Faraday cup. The measured positive ion current is the basis of relative abundance calculations.

While the refocusing properties of a magnetic field in which the ions are bent through 180° has been known for some time (Dempster 1918), it has been shown that this is a special case of the direction focusing principle and that a slightly divergent beam of charged particles is refocused if it bends through any angle less than 180°, provided it enters and leaves the magnetic field perpendicular to the boundaries of the field (Barber 1933). Stephens and Hughes (1934) carried out experiments which verified the theory using an electron beam. Figure 1

illustrates the symmetrical case of the direction focusing principle for 60°, 90° and 180° angles of deflection. A beam of ions of uniform energy which leaves the source on the focusing line is brought to focus again on the same line after being bent by the magnetic field. For all except the 180° case, the boundaries of the magnetic field are more or less sharply defined by the edges of the magnet pole faces. Since the magnetic field does extend somewhat beyond the edges, a small correction must be made in the dimensions of the pole faces in order to take care of fringe effects. Coggeshall (1947) has worked out a formula for calculating this correction.

In the single field mass spectrometer it can be shown that the fractional difference in mass which can be resolved is a function of the radius of curvature and the width of the entrance and exit slits. Resolution is therefore independent of the angle through which the ions are bent. The economy in magnet design is obvious in 90° and 60° sector type instruments since the area of the magnet pole faces is less and also the pole gap may be smaller with both source and collector systems outside the magnetic field. Certain alleged limitations of the sector types will be discussed in a later section of this paper.

2.2. *A.C. Operated Instruments*

Recent years have brought about tremendous developments in electronics and measuring techniques which are being welcomed by all who make use of modern scientific and analytical instruments. The field of mass spectrometry

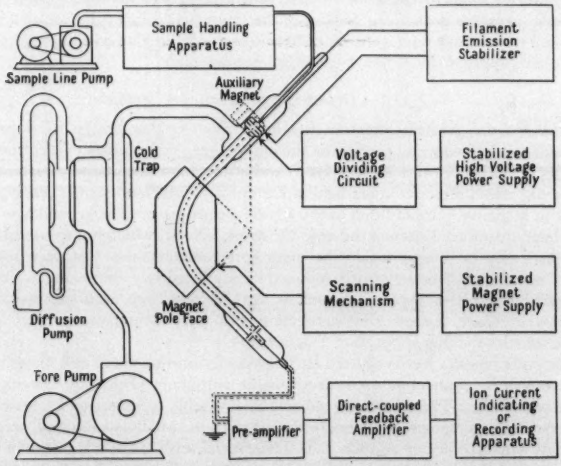


Figure 2. Block diagram of a typical A.C. operated mass spectrometer.

has shared the benefits of many new circuits and components developed during the war period, which have vastly increased the designer's scope in the field of electronically regulated power supplies. As a result, it is frequently more

economical, as well as convenient, to build a completely A.C. operated mass spectrometer than one encumbered by the unwieldy banks of batteries used almost exclusively in older instruments. Figure 2 is a block diagram of a typical A.C. operated mass spectrometer and shows the relation of the various components to the mass spectrometer tube.

Usually a hot filament provides a source of electrons which ionize the gas to be analysed. The use of batteries to heat the filament has always been unsatisfactory because even small changes in filament current produce large changes in electron emission and consequently in ion current. By the use of relatively simple electronic circuits the total emission can be held constant within 0·2% over long periods of time (Hipple, Grove and Hickam 1945). High voltage power supplies can be designed whose output voltage is constant within one part in 20,000 or better.

The magnets in earlier instruments were commonly of the low voltage type where storage batteries served as a source of power. With the advent of sector type A.C. operated instruments, high voltage–low current magnet coils came into use because the smaller current is easier to regulate electronically provided the power involved is not too large. On the other hand, many low voltage magnets are now being operated by direct current generators whose field current has been stabilized. Where extreme economy measures are necessary a permanent magnet may be used, in which case no power supply or regulating circuits are required.

Protective measures have been taken in some instruments to reduce the possibility of filament burn-out in case of a leak in the vacuum system. For example, pressure measuring devices of the ionization gauge type are now being supplied with an automatic cut-off switch to shut off the filament, high voltage or any other supply if the pressure exceeds a pre-set value.

2.3. *Ion Detecting and Measuring Methods*

Perhaps the greatest attention has been given to the design of improved methods for detecting and measuring the ion current. The balanced electrometer circuits of DuBridge and Brown (1933) and others have been largely replaced by extremely stable direct coupled feedback amplifiers which are direct reading and linear in response. In addition to the FP–54 electrometer tube, several new ones have been introduced such as the new Victoreen VX–41, which is a sub-miniature type that may be sealed inside the mass spectrometer tube. Surface leakage due to moisture is thereby greatly reduced if not completely eliminated. Several inexpensive receiving type tubes such as the 954 have been used successfully as tetrodes in space charge grid connection and have demonstrated very good characteristics for this application (Nielsen 1947).

Recently, results were reported on the use of vibrating reeds and diaphragms as inverters in connecton with electrometer amplifiers (Palevsky, Swank and Grenchik 1947). The vibrating reed acts as a variable capacitor which generates an alternating voltage proportional to the ion current, which is amplified, rectified by phase sensitive device and fed back in the usual way. The advantage of such a set-up is that a much higher degree of stability can be obtained, since it is not dependent on the D.C. stability of any of the tubes in the amplifier and is only limited by variations in contact potential of the condenser parts. These variations may be minimized by gold plating the parts and sealing them in a chamber filled with an inert gas.

At least one mass spectrometer has an ion detector consisting of a specially designed electron multiplier tube (Allen 1947) which has extremely high amplification and low background. It is claimed that single ions can be counted by this method. It is difficult to predict the usefulness of such a device in the field of mass spectrometry although at present it appears to be more suitable for counting nuclear particles.

One novel method of obtaining a continuous mass spectrum consists of superposing an alternating sweep component on the ion accelerating voltage in a conventional mass spectrometer and observing the amplified ion currents on a cathode ray oscilloscope (Forrester and Whalley 1946). In this way a panoramic view of all masses present over a considerable range can be obtained. Siri (1947) has reported a similar arrangement. Abundance measurements were said to be reliable within $2\frac{1}{2}\%$ and this was sufficient for the particular problem.

2.4. *Ion Sources*

While any of a number of arrangements may be used as a source of positive ions, the choice of one suitable for use on a mass spectrometer depends on the type of instrument and also on the particular problem. The gaseous discharge, low voltage capillary arc, and high-frequency spark are quite suitable for double

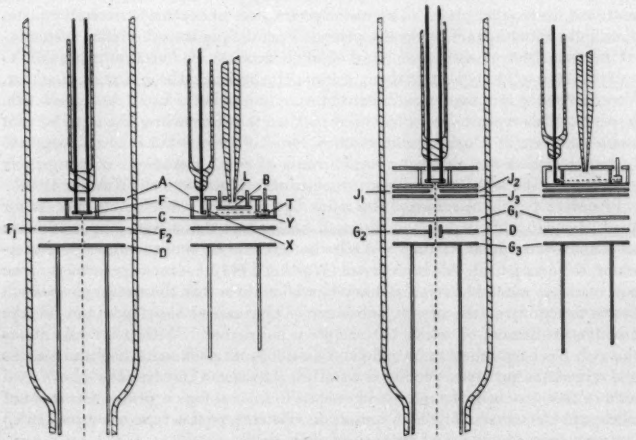


Figure 3.

(*a*) Nier type ion source. (*b*) Nier's new ion source.

focusing instruments employing photographic detection methods. In this class of instruments (mass spectrographs) some method is used to provide good focusing even though the ions have different energies. In a single focusing mass spectrometer, however, not only must the ions have a constant energy, but the intensity of the beam also must be constant. Gas mass spectrometers commonly have a type of source as shown in Figure 3(*a*) wherein the gas molecules are bombarded

by a steady electron beam of uniform energy. The electrons are emitted by a hot filament F and accelerated by about 100 volts to plate X. Some pass through the slit in the plate and travel through the ionization chamber with a uniform velocity, leaving by a slit in the plate Y on the opposite side where they are collected by a small plate usually referred to as a trap. The measured trap current gives an indication of the intensity of the electron beam in the ionization chamber. In some instruments a small magnet is used to produce a field coaxial with the beam which helps to line up the otherwise diverging electrons, thereby increasing the intensity. The gas flows from the sample tube, through the leak L, and becomes ionized upon intercepting the electron beam. A small positive potential on the repeller plate B gives the ions an initial acceleration in the direction of the slit in plate C. After passing through the slit they are accelerated by 1,000 to 2,000 volts and pass through the slit in plate D forming a steady ion beam of uniform energy which is analysed in a magnetic field according to mass.

The two half plates F_1 and F_2 are used in sector type mass spectrometers to deflect the beam a small amount in order to compensate for small errors in alignment of the slits as well as the small effect of the auxiliary magnet. A more recent type of plate system is shown in Figure 3 (*b*). It was designed by Nier (1947) in order to obtain higher ion intensities. The ions receive their initial acceleration by a drawing-out potential of 100–200 volts on plates J_1, J_2 and J_3, instead of by means of the repeller plate. The main acceleration takes place between the plates J_3 and G_1 while G_2 and D take the place of F_1 and F_2 in the other plate system.

If no suitable gaseous compound of an element can be found, it is possible to analyse some solids by heating them in a small furnace near the ionization chamber. Vapour flowing into the ionization chamber is ionized in the usual way by electron impact. This type of source has been used on instruments for the separation of stable isotopes, in which case an electron beam of several hundred milliamperes is used to ionize the vapour. Ion currents of 6 microamperes of completely resolved Ag 107 and 109 have been reported using this method (Walcher 1944).

Another type of ion source for solids has achieved considerable success for some elements. It depends on the well known fact that some compounds give off ions directly upon heating. While the mechanism is not completely understood, the general rule has been stated (Walcher 1943) that ions are formed in the case of any element whose ionization potential is lower than the energy of emission (work function) of the carrier substance. The carrier substance may be the crucible or filament on which the sample is supported. With a few exceptions this rule has been found to be valid and ions of most alkali metals, alkaline earths and rare earths have been produced directly. Lewis and Hayden (1946) have used a thermal source to make mass assignments to several fission product isotopes of rare earth elements. The high ionization efficiency of this type of source (16%) is a very great advantage where the sample is small.

Normally the ion source is operated at a high positive voltage and the ions are accelerated to earth potential. This means that the analyser tube shielding and the collector system are operated at the lowest possible impedance point, viz. earth. Ions then strike the collector with their full energy and some secondary emission is inevitable. It is, therefore, common practice to use an intermediate electron repeller plate with a fairly wide slit between the collector and the main exit slit. Usually a negative potential of a few volts is sufficient to repel any secondary electrons back to the collector. Their loss would cause an increase in

the measured current and tend to falsify the relative abundance measurements particularly if the secondary emission were not proportional to the actual ion current. This is true when different masses are collected at different accelerating voltages.

Recently Hipple and Condon (1945) noticed some rather broad peaks which appeared in the mass spectra of certain hydrocarbons. These were attributed either to ions which had lost their charge after being formed or to mass fragments which had broken up after ionization. In either case these ions should reach the collector with less energy than those which produced normal peaks. This interpretation was confirmed by an experiment wherein all ions were decelerated before collection and those which had lost more than a few volts energy were rejected. In order that the collector be operated near earth potential, the analyser tube was connected to a negative voltage and the source earthed. Under these conditions the broad peaks were found to be absent, indicating that the ions which produced them had lost energy at some point. By a study of the peak shape and the mass range in which they appeared, as well as the energy of the ions producing them, it was possible to identify the particular metastable states and measure their half lives (Hipple 1947).

2.5. *New Mass Spectrometers*

Apart from variations in the conventional type of instrument, there have been several radical departures in mass spectrometer design whose importance is difficult to predict. Numerous arrangements involving crossed electric and magnetic fields have been used, but one in particular deserves special mention. Bleakney and Hipple (1938) pointed out that the arrangement of crossed fields such that the ion paths are either prolate or curtate cycloids had focusing properties independent of the initial direction of the ions or their energy. Two models which they built confirmed the theory and demonstrated the possibilities of such a design. For a given size the curtate type was capable of twice the resolving power of the conventional 180° deflection type mass spectrometer, but involved difficulties in design. The prolate type was considered more suitable of the two. Recently a mass spectrometer of this type, called a trochotron because of the trochoidal ion paths (Monk, Graves and Horton 1947), has been built which measured 40 cm. between source and collector slits. In addition to the fact that the slits have been made very long, a widely divergent beam of ions may be used and ion currents up to 0·3 μa. have been obtained. It is to be used primarily for the separation of stable isotopes rather than for relative abundance measurements.

The mass selective properties of the linear accelerator were studied by Hartman and Smith (1939). Since then, Weisz (1946) has outlined a possible linear accelerator which is designed to make the most of these mass selective properties, and points out that with only three accelerating gaps a resolving power of one mass unit in 50 can be obtained. Harmonics are present but it is suggested that these may be distinguished if the arrival time of each pulse of ions is considered. The National Bureau of Standards (Bennett 1948) has reported the construction of a radio-frequency mass spectrometer which operates on the principles outlined by Weisz. Several types of tubes have been used in the study of negative ions where it is essential that the ion path be very short. No indication was given of the resolving power obtained. It is of interest that Stephens (1946) is now constructing a mass spectrometer based solely on the different arrival times of

different masses when a pulsed beam of ions of uniform energy travels down a long tube. Each mass produces a pulse when it strikes the collector and this is to be observed on a cathode-ray oscilloscope using a synchronized linear time-base. In this type of instrument, resolving power is not limited by slit widths but only by the length of the tube and the duration of the pulses. It is possible that a transit time mass spectrometer capable of high precision could be developed.

§3. RELATIVE ABUNDANCES OF ISOTOPES

There is a great need for mass spectrometers of high precision to determine accurately the small variations in the abundance of isotopes that occur in nature. These variations not only affect nuclear calculations and chemical atomic weights, but have other important implications in chemistry, geology and biology.

It has been known for some time that small variations did occur in the abundances of hydrogen, carbon, nitrogen and oxygen isotopes in nature. However, because of the lack of instruments of high relative precision and lack of appreciation of the importance or meaning of these small differences in isotopic content, many fields of investigation have remained unexplored.

3.1. *Thermodynamic Properties of Isotopic Substances*

Urey and Greiff (1935) showed, using the methods of statistical mechanics, that the isotopes of the light elements did have slightly different thermodynamic properties and that small fractionations of the isotopes could occur due to equilibrium processes set up in nature. These differences in thermodynamic properties depend on the fundamental vibrational frequencies, moments of inertia, zero point energies etc., of the isotopic molecules involved. Urey (1947) has calculated the isotopic exchange constants for a number of equilibrium processes and has written an excellent survey of the field to date. Of the many equilibrium constants calculated, few have been checked by experiment. For those that have been investigated, reasonably satisfactory results have been obtained. For example, the equilibrium constant at 0°c. for the reaction

$$\tfrac{1}{2}C^{16}O_{2_{gas}} + H_2{}^{18}O_{liq} \rightleftharpoons \tfrac{1}{2}C^{18}O_{2_{gas}} + H_2{}^{16}O_{liq} \qquad \ldots\ldots(1)$$

is found to be 1·046 favouring the concentration of ^{18}O in the CO_2. This agrees with the calculated value of 1·044. For this reason CO_2 oxygen is almost always 3 to 4% enriched in ^{18}O since CO_2 is usually found or prepared in equilibrium with water. Table 1 gives exchange constants for a number of isotopic reactions. Actually use was made of these small differences in thermodynamic properties of the isotopes to effect their separation. Urey and his co-workers have successfully produced concentrates of ^{13}C, ^{15}N and ^{34}S in relatively large quantities by chemical exchange methods (Thode and Urey 1939, Hutchison, Stewart and Urey 1940, Stewart and Cohen 1940).

Several surveys have been made of the isotopic content of naturally occurring carbon and oxygen samples. Dole (1935) and others found variations in the ^{18}O content of air oxygen and in natural waters of the earth using very sensitive density determinations. A similar survey has been made using mass spectrometric methods (Thode and Smith 1944). Also Nier and Gulbransen (1939) and Murphy and Nier (1941) reported small variations in the abundances of the carbon isotopes. In general, animal and vegetable carbon were found to contain less ^{13}C than limestone and igneous carbon.

Many problems arise out of this work provided instruments are available to measure these small differences with high precision. Already Kalervo Rankama (1948) has used mass spectrometer analyses of Nier to show that certain pre-Cambrian graphite deposits are probably of organic origin. The ^{13}C content of these deposits was found to be low and the isotope ratio comparable with that found for organic carbon. The earliest trace of life on the earth is thus pushed back to late Archaean times.

An equally interesting problem is the source of atmospheric oxygen. A good deal has been written on this subject, which was reviewed recently by Kamen (1946). Atmospheric oxygen is in general 3% enriched in ^{18}O over that of fresh water, and CO_2 in equilibrium with water is about 3·5% enriched. It was natural to assume therefore that atmospheric oxygen originated largely from carbon dioxide in plant photosynthesis, a process whereby plants take up CO_2 and give off

Table 1

Reaction	T (° K.)	*a* (exp.)	*a* (calc.)	Ref.
$\frac{1}{2}C^{18}O_{2_{gas}} + H_2{}^{18}O_{liq} \rightleftharpoons \frac{1}{2}C^{18}O_{2_{gas}} + H_2{}^{16}O_{liq}$	273·1	1·046	1·044	(1)
$^{15}NH_{3_{gas}} + {}^{14}NH_{4_{sol}}^{+} \rightleftharpoons {}^{14}NH_{3_{gas}} + {}^{15}NH_{4_{sol}}^{+}$	298·1	1·031 ± 0·002	1·035	(2)
$^{15}NH_{3_{gas}} + {}^{14}NH_{3_{aq}} \rightleftharpoons {}^{14}NH_{3_{gas}} + {}^{15}NH_{3_{aq}}$	298·1	1·006 ± 0·002	—	(2)
$H^{12}CN_{gas} + {}^{13}CN_{sol}^{-} \rightleftharpoons H^{13}CN_{gas} + {}^{12}CN_{sol}^{-}$	295	Near 1·026	1·030	(3)
$HC^{14}N_{gas} + C^{15}N_{sol}^{-} \rightleftharpoons HC^{15}N_{gas} + C^{14}N_{sol}^{-}$	295	Slightly >1	1·002	(4)
$^{12}CO_{3_{sol}}^{-} + {}^{13}CO_{2_{gas}} \rightleftharpoons {}^{13}CO_{3_{sol}}^{-} + {}^{12}CO_{2_{gas}}$	273·1 ?	1·017	1·016	(5)
$\frac{1}{3}C^{18}O_{3_{sol}}^{-} + H_2{}^{18}O \rightleftharpoons \frac{1}{3}C^{18}O_3{}^{--} + H_2{}^{16}O$	273·1 ?	1·036	1·033	(6)
$H^{13}CO_{3_{sol}}^{-} + {}^{13}CO_{2_{gas}} \rightleftharpoons H^{13}CO_{3_{sol}}^{-} + {}^{12}CO_{2_{gas}}$	298	1·014	—	(7)
$^{34}SO_{2_{gas}} + H^{32}SO_{3_{sol}}^{-} \rightleftharpoons {}^{32}SO_{2_{gas}} + H^{34}SO_{3_{sol}}^{-}$	298·1	1·019 ± 0·002	—	(2)
$^{36}SO_{2_{gas}} + H^{32}SO_{3_{sol}}^{-} \rightleftharpoons {}^{32}SO_{2_{gas}} + H^{36}SO_{3_{sol}}^{-}$	298·1	1·043 ± 0·004	—	(2)

(1) Weber, Wahl and Urey, 1935, *J. Chem. Phys.*, **3**, 129.
(2) Thode, Graham and Ziegler, 1945, *Can. J. Res.*, **23** B, 40.
(3) Cohen, 1940, *J. Chem. Phys.*, **8**, 588.
(4) Hutchinson, Stewart and Urey, *Ibid.*, p. 532.
(5) Nier and Gulbransen, 1939, *J. Amer. Chem. Soc.*, **61**, 697 ; Murphy and Nier, 1941, *Phys. Rev.*, **59**, 771.
(6) Dole and Slobod, 1940, *J. Amer. Chem. Soc.*, **62**, 471.
(7) Urey, Aten Jr. and Keston, 1936, *J. Chem. Phys.*, **4**, 622.

oxygen. Experiments show, however, that photosynthetic oxygen has an ^{18}O content of nearly that of fresh water and not that of the CO_2. These results lead to the paradox that air oxygen which must originate to a large extent from photosynthesis does not have the same ^{18}O content as that found for oxygen of photosynthesis. It has been suggested that exchange between oxygen and water vapour might account for this discrepancy. However, the equilibrium constants for exchange between oxygen and gaseous water at 0° and 25°c. are 1·017 and 1·014 respectively, while Dole's observed value for air oxygen is 1·03. No doubt an explanation will be found from further mass spectrometer investigations of these small variations in isotopic abundances that occur. It has been observed in our laboratory that the ^{18}O content of oxygen varies considerably depending on the source and method of production. It is our opinion that the ^{18}O content will even vary with the different oxygen groupings in a single molecule. Since proteins, starches and carbohydrates are synthesized in land plants from CO_2 and H_2O which are about 3·5% and 1% enriched in ^{18}O respectively, the ^{18}O content of a certain oxygen containing group will vary depending on the mechanism of the

synthesis; as the CO_2 is already in equilibrium with fresh water, there will be no appreciable exchange before the synthesis takes place. After synthesis of a compound, oxygen exchange between many of the groups is of course possible but will be slow and perhaps non-existent in some cases.

Other interesting problems develop from the fact that equilibrium constants have an appreciable temperature coefficient. Urey (1947) points out that geological deposits of carbon and oxygen laid down at different temperatures will have different isotopic contents. For example, considering reaction (1) above, the ^{18}O content of CO_2, and therefore of carbonates in equilibrium with water, will depend on the temperature. The equilibrium constant for the reaction varies from 1·049 at 25°c. to 1·044 at 0°c. Accurate determinations of the ^{18}O content of carbonate rocks could therefore be used to determine the temperature at which they were laid down and thus the temperature of the sea in geological time.

3.2. *Variations in Isotopic Content of Boron and Sulphur*

In spite of calculations to show differences in the thermodynamic properties of isotopes and experiments to show variations in the abundances of the isotopes of H, C and O, it has been more or less assumed that the abundances of the isotopes are constant in nature and that the small variations found are limited to H, O and C because of known chemical exchange processes. However, recent results show that these variations are much more widespread. Mass spectrometer investigations of boron samples from various geological deposits show that the ratio of ^{11}B to ^{10}B varies by as much as 3·5% (Thode, Macnamara, Lossing and Collins 1948). This is of particular interest to the physicists in view of the fact that boron is used as a standard in neutron capture cross-section measurements, and of course the standard will vary with the ^{10}B content. Still more recently similar variations in the abundances of the sulphur isotopes have been found. A considerable survey has been made at McMaster University and some correlation between geological processes and conditions of mineral formation and isotopic content has been indicated (Thode, Macnamara and Collins 1949).

In other words, we can expect to find variations in the abundances of the isotopes of many of the elements well up in the periodic table. The applications of mass spectrometry are ever increasing. As more precise instruments become available, geologists and biochemists will make use of them to determine the history and condition of formation of naturally occurring substances. The need for mass spectrometers of high relative precision is therefore very great. To meet this need, several developments in instrumentation have taken place in the past few years.

3.3. *High Precision Instruments*

In general, it is not yet possible to determine absolute abundance ratios of isotopes with an accuracy closer than ±1%. There are many reasons for this. However, by comparing samples relative to a standard under nearly identical conditions, a relative accuracy closer than ±0·1% can be obtained. Several developments have made this possible.

The older methods of measuring mass spectrometer ion currents with a balanced electrometer tube were tedious and, even more important, they were time consuming. This made necessary fairly large time corrections for drift in such factors as filament emission and sample pressure, to mention only a few.

The increasing use of highly stabilized voltage and current supplies has minimized the time corrections due to several factors, and hence increased the relative precision, but in recent years the trend has been to make use of techniques whose results are independent of changing conditions or alternatively to devise methods for reducing the time required to make the complete analysis.

There is a further reason which makes analyses of short duration important. It has been pointed out by Hönig (1945) that of the two possible types of gas flow in the mass spectrometer leak system (viscous flow and molecular flow) the ideal is molecular flow which takes place at low pressures. If the flow into the ionization chamber and the flow out of the ionization chamber are both molecular, then there will be fractionation at both places; that is, the flow will depend on the molecular weight. However, the relative densities of the components inside the ionization chamber can be shown to be independent of molecular weight and only dependent on their partial pressures in the sample reservoir. The only drawback is that because of fractionation, the relative concentration of the components is constantly changing in the reservoir. Serious errors will be introduced if the same standard

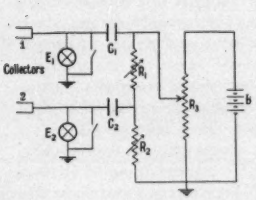


Figure 4. Dempster's method of comparing two ion currents.

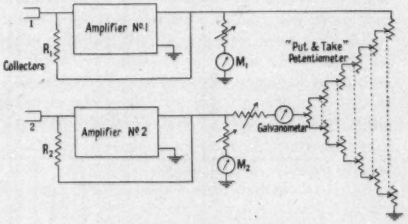


Figure 5. Nier's method of comparing two ion currents.

sample is used for long periods, particularly if the molecular weights of the components are widely different. For most isotope work the effect is small and it can be minimized if the reservoir is large and if the analysis can be completed in a short time.

The difficulties arising from changing conditions at the source of the mass spectrometer have been largely overcome in one ion current measuring technique where the ion beams for the two isotopes in question are collected at the same time and only the ratio of the currents measured. With this scheme, variations in all factors which affect both ion currents in the same proportion are unimportant since the measured isotope ratio is unchanged. Vanderberg (1941) has described such a method whereby he was able to measure the ratio of two small ion currents with an accuracy approaching one part in 10,000. The method has been applied in Dempster's laboratory (Straus 1941) to a double focusing mass spectrograph in which his Tesla spark source was subject to extreme variations. A simplified circuit is shown in Figure 4. Two ion beams are collected by individual Faraday cups and the currents used to charge two small capacitors, each of these charges being indicated on an electrometer. After a sufficient interval of time the source of ions is turned off and the ratio of the two accumulated charges is measured by the ratio of backing out voltages required to bring both electrometers back to zero.

The abundance ratio is obtained by multiplying by the ratio of the capacitances.

Nier (1947) has adopted a similar simultaneous collection scheme for comparing two ion currents using a direction focusing mass spectrometer. By this means routine abundance measurements are speeded up since isotope ratios of two samples can be compared quickly under nearly identical conditions. This makes possible a more accurate determination of the small variations that occur in isotopic ratios. Figure 5 illustrates Nier's method of measurement. Each ion current passes through a resistor and the resulting voltages are amplified by identical feedback amplifiers and compared by a "put and take" decade potentiometer system using a galvanometer as a null indicator. It is thereby possible to observe the null point continuously and note the change when a new sample is admitted. With this simultaneous collection principle, it has been possible to

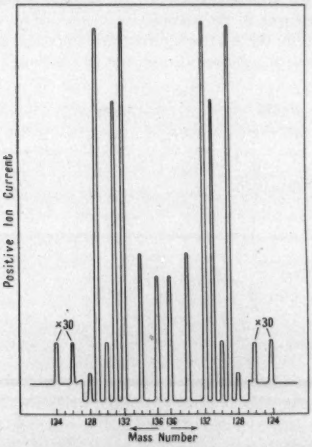


Figure 6. Double scanned record of the isotopes of normal xenon.

measure isotope abundance ratios relative to a standard to better than $\pm 0{\cdot}1\%$. For example, normal oxygen contains about $0{\cdot}200\%$ ^{18}O, and variations in this value of less than 0·0002 can be detected. This method is particularly elegant in that the measurements are independent of all factors which affect both isotopes alike and also in that the ratio is obtained directly.

The best alternative to simultaneous collection is the method which makes use of rapid automatic scanning and recording of ion currents (Smith, Lozier, Smith and Bleakney 1937, Washburn, Wiley and Rock 1943, Hipple, Grove and Hickam 1945, Inghram 1946, Lossing, Shields and Thode 1947). Commercial recorders are now being manufactured which have very fast response together with high sensitivity, properties which make them suitable for accurately recording the output signal of the electrometer amplifier. Figure 6 shows a double scanned record of the nine isotopes of normal xenon. A statistical study of a number of recorded spectrograms obtained over a very short time interval makes possible a

precision comparable with that obtained by the simultaneous collection method. The advantage of this method is at once apparent when it is realized that in one operation the relative abundance of all nine isotopes can be measured with a relative precision closer than 0·1%.

Scanning is usually accomplished by the use of a driven helical potentiometer to vary the magnet current or the accelerating voltage, whichever is preferred. Recently a new type of scanning device has been patented (Olesen 1946) which utilizes an R–C component in the grid circuit of a vacuum tube in order to vary the accelerating voltage at a rate suitable for scanning. It is quite conceivable that a similar device could be used for magnetic scanning.

Quite often the operation of an automatic recording instrument is enhanced if means are provided for automatically pre-selecting an appropriate recorder sensitivity so that the recorded peak always occupies a considerable portion of the scale in order to minimize measuring errors. Several methods have been described for doing this. Washburn, Wiley and Rock (1943) used four galvanometers of different sensitivity which recorded simultaneously on photographic paper. Grove and Hipple (1947) use a system which makes two records consecutively, one on a logarithmic sensitivity scale and one on a linear scale using the sensitivity which was automatically selected during the first scan. A shunt selecting unit has been described (Lossing, Shields and Thode 1947) which is entirely electronic in operation and which requires only one scanning operation with each peak being recorded at the proper level.

Another useful device has been described by Hipple, Grove and Hickam (1945) which provides means for marking the mass number on the recorder chart directly below the peaks. The mass marker is controlled by the voltage output of a small rotating armature in the mass spectrometer magnetic field.

§4. ABSOLUTE ABUNDANCE MEASUREMENTS

Presumably if all sources of discrimination could be positively eliminated, the abundance data obtained on a mass spectrometer might be taken as absolute. In general this is not completely possible and the widely different values of different investigators serve to bear out the fact. Discrimination may be of two kinds: mass discrimination, in which the efficiency of production and collection of ions is different for different masses; and peak height discrimination, which is caused by varying sensitivity with intensity. Peak height discrimination is the least serious of the two and is caused mainly by non linear components in the current measuring equipment. The use of properly designed feedback amplifiers precludes the possibility of non-linearity from this source and indicating devices may be obtained which are linear within the limits of experimental error. The grid leak in the pre-amplifier circuit is the most common source of trouble. Manufacturers of grid leak resistors usually caution against the use of voltages greater than about 15 volts across them. The resistance is claimed to be essentially constant for lower voltages. Some resistors tend to become polarized when high currents are passed and retain a small internal source of E.M.F. for a short time afterwards. It is often necessary to hand-pick grid resistors and select one which does not exhibit abnormalities of this kind.

Mass discrimination is present wherever the various mass components of the sample display different properties in any process in the mass spectrometer. These processes include gas flow in the sample line, ionization, acceleration and

collection. With regard to gas flow, it was pointed out earlier that molecular flow is ideal from the standpoint of maintaining the same relative concentration on each component in the ionization chamber as that in the sample tube. Nevertheless, a capillary leak is quite commonly used for isotope work where the differences in molecular weight are small. Somewhat higher pressures are used behind the leak in this case, but discrimination may still be almost negligible. The existence of discrimination may be detected by noting the difference in the measured ratios at different pressures and if it is found to be appreciable, a correction may be made by extrapolating to zero pressure.

In the ion source arrangement there are a large number of related effects, some of which are selective with regard to mass. For example, the position of the electron beam may change slightly as the accelerating voltage or magnetic field is varied to collect different masses, or the ion paths in the source assembly may be different for different masses and thus be subject to different conditions. Several of these effects have been calculated (Jordan and Coggeshall 1942) and the results led to the earlier contention that the most important criterion for an instrument free of mass discrimination was that the individual ion paths in the source should be the same for each mass under the conditions of collection. The method of simultaneous collection has been criticized on the basis that different masses originate at different points in the electron beam and in general the method is not suitable for making absolute measurements. Also sector types are open to criticism for the same reason. Nier, in reporting his first instrument of this type, cautioned that it was not free from mass discrimination, but pointed out its suitability for making relative measurements. Sector instruments in particular exhibit much higher collection efficiencies at high accelerating voltages. Thus, if the accelerating voltage is varied to permit collection of different masses (electrostatic scanning) the lighter mass is collected under more efficient conditions and the error may be several per cent. Coggeshall (1944) has attributed this variation in efficiency to the initial thermal energy of the molecules at the moment they become ionized. By a theoretical discussion of the conditions in a magnetic field free source, he has shown that a temperature of about 500°C. produces discrimination of about the same order of magnitude as that observed experimentally. He says that sector instruments should be free from discrimination of this kind if magnetic scanning is used. The ion paths will still be different for different masses in this case, and it would therefore appear that neither method of scanning is free from discrimination. However, it is a question of which effect is the greater. Washburn and Berry (1946) have expressed the opinion that the choice of a scanning method is merely a matter of convenience, but since they are interested primarily in hydrocarbon analysis, considerable mass discrimination may be found using either method.

A good indication of the extent of mass discrimination obtained under various conditions can be obtained experimentally. A gaseous molecule such as BF_3, will give rise to the following singly charged ions in a mass spectrometer: $^{10}B^+$, $^{11}B^+$; $^{10}BF^+$, $^{11}BF^+$; $^{10}BF_2^+$ $^{11}BF_2^+$; and $^{10}BF_3^+$, $^{11}BF_3^+$. The isotopic ratio of ^{11}B to ^{10}B can therefore be determined by comparing the isotopic ion currents in any of these mass ranges. If mass discrimination is serious, then the most reliable isotopic ratio will be obtained in the higher mass range where the percentage mass difference is the least, whereas if there is no appreciable mass discrimination then all determinations should compare and be equally reliable. Inghram (1946) used

a 60° deflection mass spectrometer to determine the isotopic constitution of boron and silicon. The ratios of $^{11}B/^{10}B$ obtained were within $\pm 1\%$ regardless of the mass range of the ions, provided magnetic scanning of the ion currents was used. Electrostatic scanning on the other hand gave rise to mass discrimination effects amounting to $\pm 20\%$ for a 10% mass difference.

Thode, Macnamara, Lossing and Collins (1948) report similar results in so far as mass discrimination effects are concerned. However, these mass discrimination effects due to electrostatic scanning were found to be less for a 180° deflection instrument than for a 90° type. This might be expected since the two collimating slits between which the ions are accelerated in the ion source are in general closer together in a 180° instrument, and according to Coggeshall's arguments, mass discrimination arising from the initial energies of the ions should be less in this case.

Since the latter workers found the ratio of ^{11}B to ^{10}B to vary considerably depending on the geological source of boron, it is not possible to compare results obtained in different laboratories unless the history of the samples is known.

There are further factors which can produce abnormalities. Not the least of these is the effect of space charge conditions in the source. With a fairly low electron beam intensity very little difficulty may be encountered, but if the ratios are found to depend at all on electron intensity, the usual practice is to extrapolate to zero trap current. When solids are to be used in an ion source, differences in vapour pressure are important. In sources where the solid emits ions directly upon heating, differences in ionization potential may be a source of discrimination. For example, Hundley (1927) found that the discrepancies in the measured abundances of ^{6}Li and ^{7}Li could be accounted for by the difference in their ionization potentials. As a final practical consideration, it is extremely important that all surfaces in the mass spectrometer be scrupulously clean. Charges tend to build up on any insulating dirt film, and while it is impossible to calculate their effect, experience has shown that a dirty tube is often responsible for asymmetrical peaks and many other irregularities.

In general all types of mass discrimination are more serious in the low mass range where the fractional mass difference is large. For the large volume of isotope work, these differences may be very small, but they can never be disregarded. In addition to the known sources of mass discrimination, there are most certainly others, since it is a rare occasion when results taken on different days agree to within 0·5% even when precautions are taken to eliminate all known sources of discrimination. In view of this fact, it will be easily realized that since absolute abundance measurements of high accuracy are essential to some problems, modern precision instruments still leave something to be desired in this respect.

§5. NUCLEAR PROBLEMS

In the fission of uranium, many isotopes of the fission elements are produced. The primary products have more neutrons for their charge than the stable nuclei and undergo β-disintegrations in several steps ending in stable nuclear structures. Some sixty-four such fission chains are now known (Plutonium Project Report 1946).

It was clear from the first fission product studies that the mass spectrometer would eventually be useful in separation problems, mass identification, and isotope abundance measurements. Actually the mass spectrometer provides the only convenient means for identifying stable isotopes which are end products of

fission chains. However, the early work on fission products involved such small quantities of material that only radio-chemical methods were considered sensitive enough to detect and follow the radioactive isotopes. During the war, techniques were developed for extracting and purifying amounts of fission gas of the order of a cubic millimetre, and these amounts have been sufficient for mass spectrometer investigations (Arrol, Chackett and Epstein 1947). Thode and Graham (1947) reported complete analysis of 1 mm^3 samples of fission gas. With these samples it was possible to detect 10^9 atoms of an isotopic constituent. It should be possible to extend this limit further, since only about 20% of these small samples actually pass through the mass spectrometer, the remainder being available for further experiments. These experiments with small fission gas samples indicate the micro nature of mass spectrometer analyses.

Because of the large number of stable isotopes of xenon and krypton (see Figures 7 and 8) which block further β-disintegrations of fission product chains, the mass spectrometer investigations of fission gas samples give information about a relatively large number of these chains.

In these investigations the irradiated uranium samples are allowed to stand for several months before extraction of the fission gases, to permit the decay of most fission product chains to stable isotopes. Figures 9 and 10 show recent spectrograms of the isotopes of xenon and krypton identified in fission by mass spectrometry, and their distribution as compared to the abundances of those identified in nature (Thode, Macnamara and Collins 1948).

The marked difference in the number of isotopes present and their abundances is, of course, apparent. By a comparison of the mass spectrometer abundance data with the fission yield data for known chains ending in xenon, it was possible to assign masses to the respective chains with certainty. These are:—

^{129}Sb $\xrightarrow{\beta}$	^{129}Te $\xrightarrow{\beta}$	^{129}I $\xrightarrow{\beta}$	^{129}Xe
4·2 hr.	32 dy. 72 min.	not found	not found
	^{131}Te $\xrightarrow{\beta}$	^{131}I $\xrightarrow{\beta}$	^{131}Xe
	30 min.	8 day	stable
^{132}Sb $\xrightarrow{\beta}$	^{132}Te $\xrightarrow{\beta}$	^{132}I $\xrightarrow{\beta}$	^{132}Xe
5 min.	77 hr.	2·4 hr.	stable
^{134}Sb $\xrightarrow{\beta}$	^{134}Te $\xrightarrow{\beta}$	^{134}I $\longrightarrow$	^{134}Xe
10 min.	43 min.	54 min.	stable
		^{136}I $\xrightarrow{\beta}$	^{136}Xe
		1·8 min.	stable

In the unravelling of these fission chains the mass spectrometer therefore made possible not only the identification of stable isotopes, but the mass assignment of the corresponding fission chains.

The isotope ^{85}Kr which does not appear in naturally occurring krypton is radioactive. It is present in all fission product krypton samples but its relative concentration is less in older samples. The existence of this isotope was not suspected previous to the mass spectrometer investigations because a 4 hour ^{85}Kr was already shown to exist and this isotope would have completely decayed in samples six months to a year old. The ^{85}Kr found is a long lived isomer of the 4 hour ^{85}Kr. A long lived isotope of krypton, of unknown mass, was also found by counting experiments. Its half life of 9·4 years was measured directly by the mass spectrometer method discussed below.

^{129}Xe is not present in fission product samples several years old, at least not in amounts within the limits of detection, i.e. 1 part in 40,000. However, the fission yield for the mass 129 chain is quite considerable. This suggests the existence of a long lived ^{129}I nucleus which up to the present has escaped detection. The mass spectrometer could be used to identify ^{129}I directly.

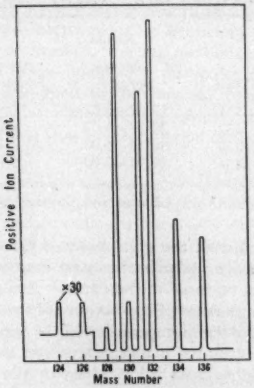


Figure 7. Isotopes of normal xenon.

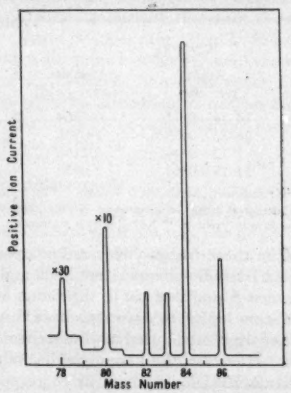


Figure 8. Isotopes of normal krypton.

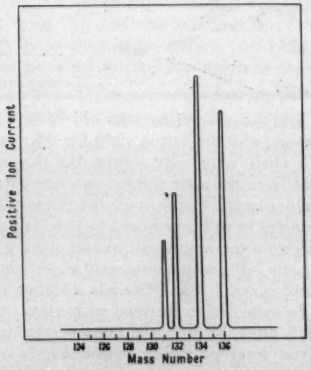


Figure 9. Isotopes of fission product xenon.

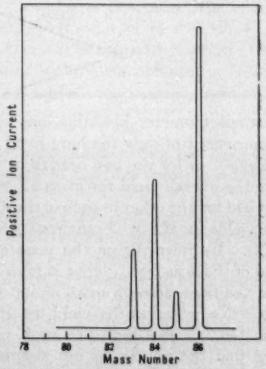


Figure 10. Isotopes of fission product krypton.

5.1. *Fission Yields by the Mass Spectrometer Method*

The relative abundances of stable isotopes of a given element resulting from the decay of fission chains give direct information about the fission yields for

these mass chains. In this regard, the fission gases offer a very fruitful field of investigation because of the large number of fission chains which end in either krypton or xenon.

Table 2 gives typical isotope abundance data for fission product xenon and krypton, together with absolute fission yields of corresponding mass chains.

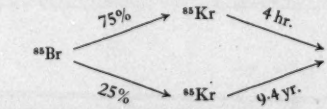
Table 2

Krypton			Xenon		
Mass	Atom per cent	Absolute fission yield†	Mass	Atom per cent	Absolute fission yield†
83	14·34−0·02	0·5864	131	13·62−0·02	2·8
84	26·72−0·03	1·0926	132	20·30−0·02	4·1667
*85	7·72−0·06	0·2168	134	36·13−0·03	7·41
86	51·19−0·10	2·0936	136	29·94−0·03	6·14

* ^{85}Kr radioactive half life 9·4 years. Percentage obtained by extrapolation to zero time.

† Calculated using experimental Xe/Kr ratio=5·02 (Arrol, Chackett and Epstein 1947) and absolute fission yield of ^{131}Xe=2·8 as a standard.

When these fission yields are compared with the great body of yield data obtained by radio-chemical methods and as compiled in the mass-fission yield curve, it is found that six of the values can be fitted to the curve. Actually it would seem logical to draw the curve through these points because of the direct nature of the method used to obtain them and the high precision of the measurements. However, the fission yield obtained by the mass spectrometer method for ^{85}Kr is 75% lower than the corresponding mass values given by the mass fission yield curve. This discrepancy is easily explained as there are two isomers of ^{85}Kr which decay independently according to the mechanism



.....(2)

Mass spectrometer investigations of samples several months old involve the measurement of only the long lived isomer and the fission yield for ^{85}Kr given in Table 2 is for the one branch of the chain only. By subtracting this yield from the overall yield for mass 85, read from the mass yield curve, one obtains the yield for the other branch of the chain, namely, that through the 4 hour ^{85}Kr. The yields for the two branches are found to be in the ratio of 3 to 1 as indicated in (2). By normalizing the mass spectrometer abundance values, taking the yield of ^{131}Xe as 2·8%, ^{134}Xe is found to be 7·4% which obviously would not be expected from smooth mass fission yield curve. Also ^{134}Xe has a higher yield than ^{136}Xe, whereas the yield would be expected to increase with mass up to mass 140. These high precision yields obtained with the mass spectrometer show that the process is not simply the decay of fission chains, but is more complicated than was previously thought.

An important feature of the mass spectrometer method for determining fission yields is the high accuracy and particularly the high precision obtainable. It is possible with automatic recording equipment to determine isotope abundances with an absolute accuracy of 1% and with a precision of ±0·1%. This is to be compared with errors of ±10% in the values obtained by radio-chemical methods. The high internal precision obtainable with the mass spectrometer

makes it possible to compare samples extracted from uranium irradiated under different conditions and measure small significant differences in the distribution of fission product isotopes. Actually, one would expect fine structure in the curve for thermal fission yield of ^{235}U products as a function of mass number because of delayed neutron emission, coincidental emission, or irregularities in stability of primary fission products. It seems important, therefore, to extend the mass spectrometer investigations to fission isotopes other than krypton and xenon, as standard chemical methods are not adequate to show up small variations in the fission yield of a particular chain. However, the mass spectrometer method applied to the rare gas fission isotopes already provides an elegant procedure for investigating quickly the fission yields of eight isotopes produced by fast and slow neutron fission of the fissile elements.

5.2. *Half Life Measurements*

The mass spectrometer provides a method for the determination of the half life of certain radioactive isotopes where standard methods may not be adequate because of contamination from other active constituents or because of difficulty in determining specific activity. For example, the half life of ^{85}Kr has been accurately determined mass spectrometrically. The method involves the determination of the concentration of ^{85}Kr relative to a stable isotope and following the decrease in this concentration with time. The high precision possible in the determination of relative abundances (0·1% or better) makes reasonably accurate half life measurements attainable over a considerable range of activities.

In the determination of the half life of ^{14}C, mass spectrometer abundance data on the isotopes of carbon is necessary in order to determine the specific activity of ^{14}C. Norris and Inghram (1948) obtained a value of 5,100 ± 200 years by this method, whereas Hawkings, Hunter, Mann and Stevens (1948) using the same method but different counting techniques obtained a value of 6,400 ± 200 years.

5.3. *Assignment of Mass to Fission Isotopes*

Considerable mass spectrograph and mass spectrometer work has been done with the non-gaseous fission products, particularly in regard to mass identification of radioactive fission isotopes.

Hayden and Inghram (1946) have reported a new mass spectrograph technique used for the assignment of mass to 46 hour samarium and 9·2 hour europium. Their prodecure for the establishment of mass 153 to the 46 hour samarium isotope is as follows: a solution of Sm_2O_3 is first irradiated with slow neutrons to produce the 46 hour samarium activity, an activity also present in fission product material. Part of this activity is placed on a filament source of a mass spectrograph. A mass record is then obtained, but before development, the photographic plate is placed in contact with an unexposed plate and left for a suitable period of time. After development, the original plate shows the normal spectrum of samarium while the second plate shows only two lines. These lines correspond to masses 153 and 169 on the original plate and represent "radio autographs" of the radioactive isotopes of these masses. The two lines are both due to ^{153}Sm since SmO^+ ions are produced as well as Sm^+ ions. The mass of 46 hour samarium is therefore 153. In a similar manner the mass of 9·2 hour europium was established as 152.

Rall (1946) investigated Pd and Ir which had been irradiated with slow neutrons in a double focusing mass spectrograph using a high-frequency spark to produce Pd and Ir ions. Using the radio-autograph technique, the masses of 13 hour Pd, 60 day Ir and 19 hour Ir were established as 109, 192 and 194 respectively. According to Rall, approximately 10^5 beta particles are required to produce a line on the transfer plate as compared with 10^8 ions of 12,000 volts energy.

Later Lewis and Hayden (1946) described a mass spectrograph for fission product studies which makes use of the thermal emission of ions from solid samples. According to their results, ion emission for the rare earth elements is very good. Further, the emission obtained is characteristic of the element and in some cases the element can be established by the type of emission obtained. In these cases not only the mass of a radioactive isotope can be established, but the element as well. The position of the radioactive isotopes on the spectrograph plate may be located by means of a Geiger counter or by means of a "radio-autograph" of the spectrograph plate. This instrument was used primarily for the study of isotopes of rare earth elements. The masses of the fission isotopes 55 day strontium, 30 year strontium, 57 day yttrium, 1 year ruthenium, 35 year caesium, 40 hour lathanium, 25 day cerium, 14 day praesiodymium, 275 day cerium, 1–2 year illinium, long lived samarium, and 2–3 year europium were found to be 89, 91, 106, 137, 140, 141, 143, 144, 147, 151 and 155 respectively.

Also Inghram, Hess and Hayden (1947), using a Nier type mass spectrometer with a vibrating reed electrometer and pen recorder, established the mass number of a long lived isotope of element 43, technetium, previously discovered in fission. The sample of fission technetium used for their work was distilled into the ionization chamber of the mass spectrometer as NH_4TcO_4. The mass spectrograms obtained covered the mass range 98 to 103. In this range a pattern for doubly charged mercury ions is obtained. However, with technetium distilling into the ion source, the ion current corresponding to mass 99 more than doubled indicating the presence of ^{99}Tc isotope.

The further extension of the mass spectrometer studies to fission products other than krypton and xenon offers a fruitful field for investigation. However, because of the elegance of techniques for extracting and purifying fission gases and ease of handling rare gases in a mass spectrometer, the fission gases offer the best opportunity for investigation the distribution of fission products obtained under different conditions of uranium irradiation etc. Work is now in progress at McMaster University to determine the yields of the krypton and xenon isotopes produced in fission as a function of neutron energy, neutron flux and fissile element.

REFERENCES

ALLEN, J. S., 1947, *Rev. Sci. Instrum.*, **18,** 739.

ARROL, W. J., CHACKETT, K. F., and EPSTEIN, S., 1947, *Proceedings of Conference on Nuclear Chemistry* (Hamilton, Ont.: McMaster Univ.), p. 108.

BAINBRIDGE, K. T., and JORDAN, E. B., 1936, *Phys. Rev.*, **49,** 421.

BARBER, N. F., 1933, *Proc. Leeds Phil. Soc.*, **2,** 427.

BENNETT, W. H., 1948, *Radio Electronic-Engineering*, **40,** No. 5, 10.

BLEAKNEY, W., and HIPPLE, J. A., Jr., 1938, *Phys. Rev.*, **53,** 521.

COGGESHALL, N. D., 1944, *J. Chem. Phys.*, **12,** 19; 1947, *Phys. Rev.*, **71,** 482.

DEMPSTER, A. J., 1918, *Phys. Rev.*, **11,** 316; 1935, *Proc. Amer. Phil. Soc.*, **75,** 755.

DOLE, M., 1935, *Jour. Amer. Chem. Soc.*, **57**, 2731.
DUBRIDGE, L. A., and BROWN, H., 1933, *Rev. Sci. Instrum.*, **4**, 532.
FORRESTER, A. T., and WHALLEY, W. B., 1946, *Rev. Sci. Instrum.*, **17**, 549.
FOX, R. E., HIPPLE, J. A., Jr., and WILLIAMS, T. W., 1944, *Phys. Rev.*, **65**, 353.
GRAHAM, R. L., HARKNESS, A. L., and THODE, H. G., 1947, *J. Sci. Instrum.*, **24**, 119.
GROVE, D. J., and HIPPLE, J. A., Jr., 1947, *Rev. Sci. Instrum.*, **18**, 837.
HARTMAN, P. L., and SMITH, L. P., 1939, *Rev. Sci. Instrum.*, **10**, 223.
HAWKINGS, R. C., HUNTER, R. F., MANN, W. B., and STEPHENS, W. H., 1948, *Phys. Rev.*, **74**, 696.
HAYDEN, R. J., and INGHRAM, M. G., 1946, *Phys. Rev.*, **70**, 89.
HIPPLE, J. A., Jr., 1947, *Phys. Rev.*, **71**, 594.
HIPPLE, J. A., Jr., and CONDON, E. U., 1945, *Phys. Rev.*, **68**, 54.
HIPPLE, J. A., Jr., GROVE, D. J., and HICKAM, W. M., 1945, *Rev. Sci. Instrum.*, **16**, 69.
HÖNIG, R. E., 1945, *J. Appl. Phys.*, **16**, 646.
HUNDLEY, J. L., 1927, *Phys. Rev.*, **30**, 864.
HUTCHISON, C. A., STEWART, D. W., and UREY, H. C., 1940, *J. Chem. Phys.*, **8**, 532.
INGHRAM, M. G., 1946, *Phys. Rev.*, **70**, 653.
INGHRAM, M. G., HESS, D. C., and HAYDEN, R. J., 1947, *Phys. Rev.*, **72**, 1269.
JORDAN, E. B., 1940, *Phys. Rev.*, **57**, 1072.
JORDAN, E. B., and COGGESHALL, N. D., 1942, *J. Appl. Phys.*, **13**, 539.
KAMEN, M., 1946, *Bull. Amer. Nat. Hist., New York*, 87.
LEWIS, L. G., and HAYDEN, R. J., 1946, *Phys. Rev.*, **70**, 111.
LOSSING, F. P., SHIELDS, R. B., and THODE, H. G., 1947, *Canad. J. Res.*, **25** B, 397.
MATTAUCH, J., 1936, *Phys. Rev.*, **50**, 617.
MONK, G. W., GRAVES, J. D., and HORTON, J. L., 1947, *Rev. Sci. Instrum.*, **18**, 796.
MURPHY, B. F., and NIER, A. O., 1941, *Phys. Rev.*, **59**, 771.
NIELSEN, C. E., 1947, *Rev. Sci. Instrum.*, **18**, 18.
NIER, A. O., 1947, *Rev. Sci. Instrum.*, **18**, 398.
NIER, A. O., and GULBRANSEN, E. A., 1939, *J. Amer. Chem. Soc.*, **61**, 697.
NORRIS, L. D., and INGHRAM, M. G., 1948, *Phys. Rev.*, **73**, 350.
OLESEN, R. C., 1946, *Report to Consolidated Engineering Corp.*, U.S. Patent No. 2397560.
PALEVSKY, H., SWANK, R. K., and GRENCHIK, R., 1947, *Rev. Sci. Instrum.*, **18**, 298.
PLUTONIUM PROJECT REPORT, 1946, *J. Amer. Chem. Soc.*, **68**, 2411.
RALL, W., 1946, *Phys. Rev.*, **70**, 112A.
RANKAMA, K., 1948, *Bull. Geol. Soc. Amer.*, **59**, 389.
SHAW, A. E., and RALL, W., 1947, *Rev. Sci. Instrum.*, **18**, 278.
SIRI, W., 1947, *Rev. Sci. Instrum.*, **18**, 540.
SMITH, P. T., LOZIER, W. W., SMITH, L. G., and BLEAKNEY, W., 1937, *Rev. Sci. Instrum.*, **8**, 51.
STEPHENS, W. E., 1934, *Phys. Rev.*, **45**, 513 ; 1946, *Ibid.*, **67**, 691.
STEPHENS, W. E., and HUGHES, A. L., 1934, *Phys. Rev.*, **45**, 123A.
STEWART, D. W., and COHEN, K., 1940, *J. Chem. Phys.*, **8**, 904.
STRAUS, H. A., 1941, *Phys. Rev.*, **59**, 430.
THODE, H. G., and GRAHAM, R. L., 1947, *Canad. J. Res.*, A **25**, 1–14.
THODE, H. G., MACNAMARA, J., and COLLINS, C. B., 1949, *Canad. J. Res.*, April.
THODE, H. G., MACNAMARA, J., LOSSING, F. P., and COLLINS, C. B., 1948, *J. Amer. Chem. Soc.*, **70**, 3008.
THODE, H. G., and SMITH, S. R., 1944, *Montreal Report M.C.*, 57.
THODE, H. G., and UREY, H. C., 1939, *J. Chem. Phys.*, **7**, 34.
UREY, H. C., 1947, *J. Chem. Soc.*, **8**, 562.
UREY, H. C., and GREIFF, L., 1935, *J. Amer. Chem. Soc.*, **57**, 321.
VANDERBERG, R. M., 1941, *Phys. Rev.*, **59**, 114 A.
WALCHER, W., 1943, *Z. Phys.*, **121**, 604 ; 1944, *Ibid.*, **122**, 62.
WASHBURN, H. W., and BERRY, C. E., 1946, *Phys. Rev.*, **70**, 559.
WASHBURN, H. W., WILEY, H. F., and ROCK, S. M., 1943, *Ind. Eng. Chem., Anal. Ed.*, **15**, 541.
WEISZ, P. B., 1946, *Phys. Rev.*, **70**, 91.

NUCLEAR PARAMAGNETISM

By B. V. ROLLIN
The Clarendon Laboratory, Oxford

§1. INTRODUCTION

ALTHOUGH nuclear magnetism has been extensively investigated by the methods of spectroscopy and by experiments with atomic and molecular beams, it is only in recent years that it has become possible to detect magnetic effects in ordinary matter due to the magnetic moments of the constituent nuclei. The main difficulty is that nuclear moments are only of the order of a thousandth of the Bohr magneton, so that the susceptibility of a substance, due to nuclear magnetism, is about a million times smaller than for electron paramagnetism. If there are N nuclei with spin $I\hbar/2\pi$ and moment μ the susceptibility χ_0 may be calculated from the ideal Curie relation:

$$\chi_0 = \frac{1}{3}\frac{I+1}{I}\frac{N\mu^2}{kT}.$$

For the proton $I=\frac{1}{2}$, $\mu = 1{\cdot}4\times10^{-23}$, so that the susceptibility calculated for 1 cm³ of water at room temperature is $\chi_0 = 3{\cdot}4\times10^{-10}$.

By making measurements at very low temperatures it is possible to observe the nuclear susceptibility by ordinary methods; this has been done by Lasarew and Schubnikow (1937) in experiments on solid ortho hydrogen at 2° K., but such measurements are extremely difficult to carry out. Recently, new methods for detection and measurement of nuclear paramagnetism have been employed, based on the principle of radio-frequency resonance which was first suggested by C. J. Gorter (1936). These methods, which were developed independently by Purcell, Torrey and Pound (1946) and by Bloch, Hansen and Packard (1946) have opened a new field of research and it is the object of this Report to give a brief review of the progress which has been made in this work up to the present time.

§2. NUCLEAR MAGNETIC RESONANCE

The radio-frequency resonance method, in its application to nuclear paramagnetism, has several features in common with the well known experiments of Rabi and his collaborators on atomic and molecular beams. The basic principle is the detection of the transitions of a nucleus in a strong constant magnetic field due to the application of a small resonant oscillatory magnetic field. For a nucleus of spin I in a magnetic field H_0 there are $2I+1$ energy states corresponding to the possible orientations in the field. The energy difference between the states is $\mu H_0/I$ and transitions can be induced by a transverse oscillating, or rotating, field of frequency ν_0 given by the resonance relation:

$$h\nu_0 = \mu H_0/I. \qquad \ldots\ldots(1)$$

In the method of Rabi the transitions are observed by their effect on the deflection of molecules in an inhomogeneous field. In experiments on nuclear paramagnetism, where a large number of nuclei are influenced simultaneously, the transitions can be observed by the absorption of energy from the oscillatory

field or by other effects depending on the transverse magnetization which is produced in the specimen. For practical values of the field H_0, the frequency required for nuclear resonance is of the order of a few megacycles per second.

In the original experiments of Gorter (1936) an attempt was made to detect resonant absorption by a calorimetric method. The measurements were made at 14° K. and a search was made for heating of the specimen due to resonant absorption by the protons in aluminium potassium alum and by ^{7}Li in lithium fluoride, but the method was not sufficiently sensitive to enable positive results to be obtained. A subsequent experiment by Gorter and Broer (1942) was designed to detect the dispersion, or change of transverse susceptibility, in the neighbourhood of resonance, by observing the change of frequency of the oscillator which supplied the radio-frequency field. Although a sensitive beat frequency method was used, no effect could be found.

The methods which have been developed recently are, in some respects, similar to that used by Zavoisky (1945) for investigation of electron paramagnetism, with the exception of course, that, in order to detect nuclear effects, it is necessary to use much more refined apparatus. In the method devised by Purcell, Torrey and Pound (1946) the resonance is observed by the damping effect of the nuclear absorption on the tuned circuit which provides the radio-frequency field. In the experiments of Bloch, Hansen and Packard (1946) the resonance is detected by the voltage induced in a separate tuned circuit by the transverse nuclear magnetization. Before describing the experimental arrangement used in these methods it will be convenient to consider the nature and magnitude of the resonance effects which may be expected under various conditions.

§ 3. NATURE AND MAGNITUDE OF THE RESONANCE EFFECTS

The resonance effects can be induced either by an oscillatory field or by a rotating field which has the same direction of rotation as the Larmor precession of the nuclei. An oscillating field $2H_1 \sin \omega t$ can be considered to consist of two rotating components of amplitude H_1 rotating in opposite directions; only the component which has the direction of rotation of the Larmor precession is effective, so that the transverse magnetization which is induced in the neighbourhood of resonance is rotating and not oscillatory.

The magnitude of the transverse magnetization depends on a number of factors such as the values of H_0 and H_1, the spin–spin and spin–lattice interactions and the rate of attainment of resonance. Calculations of the resonance effects can be made either from the quantum theory of absorption and dispersion or, by finding the effects of the fields H_0 and H_1 on the total magnetization of the specimen, from the equations of classical dynamics.

In the quantum method (Gorter 1936, Gorter and Kronig 1936, Gorter and Broer 1942, Frenkel 1945, Purcell, Torrey and Pound 1946, Bloembergen, Purcell and Pound 1948, Bloembergen 1948a) the probability of absorptio nor emission of a quantum by a nucleus, under the influence of the radio-frequency field, is obtained from the transition probability for magnetic dipole transitions. For $I = \frac{1}{2}$ the transition probability is simply the Einstein B coefficient. To find the resultant effect, it is necessary to sum over all nuclei taking account of the distribution over the various levels and also of the broadening and spreading of the levels due to mutual interaction of the nuclei and magnetic field inhomogeneities. The probabilities of absorption and stimulated emission are equal,

so that there would be no resultant absorption if all the levels were equally populated. However, if the radio-frequency field is very weak, and the nuclei are in thermal equilibrium, there is a Boltzmann distribution over the levels and an absorption proportional to the level population.

If the transverse magnetization is resolved into the components U and V, where U is in the direction of the rotating field H_1, and V is at right angles (Figure 3 (*b*)), we may write $U=\chi' H_1$ and $V=\chi'' H_1$, where χ' may be described as the transverse susceptibility and χ'' the absorption coefficient. It may be shown that, if H_1 is very small, the maximum values obtained for χ' and χ'' under equilibrium conditions are given by

$$\chi' = \tfrac{1}{2}\left(\frac{\nu}{\Delta\nu}\right)\chi_0 = \tfrac{1}{2}\left(\frac{H_0}{\Delta H}\right)\chi_0; \qquad \chi'' = \left(\frac{\nu}{\Delta\nu}\right)\chi_0 = \left(\frac{H_0}{\Delta H}\right)\chi_0$$

where $\Delta\nu$ and ΔH are the width of the resonance expressed in terms of frequency or magnetic field. The variation of χ' and χ'' with frequency is shown in Figure 1; whereas χ'' has a maximum on exact resonance, χ' is maximum at a frequency $\Delta\nu$ from exact resonance. For water at room temperature with $H_0=2{,}000$ gauss, the maximum values of χ' and χ'' are of the order of 10^{-5} per cm^3.

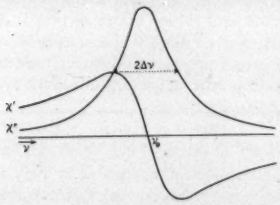


Figure 1. Variation of the transverse susceptibility χ' and the absorption coefficient χ'' with frequency in the neighbourhood of resonance.

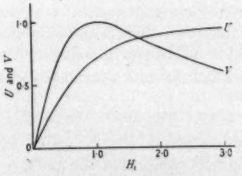


Figure 2. Variation of the components of transverse magnetization U and V with strength of radio-frequency field H_1. The unit of H_1 is $1/\gamma(T_1T_2)^{\frac{1}{2}}$ and the unit of U and V, $\frac{1}{2}(T_2/T_1)^{\frac{1}{2}}M_0$. ($T_2=1/2\pi\Delta\nu$.)

The effect of a strong radio-frequency field is to tend to equalize the population of the levels, so that the magnetization of the specimen is reduced and the values of χ' and χ'' are lower than when H_1 is small. This effect may be approximately described as an elevation of the temperature of the nuclear spin system, due to absorption of energy from the field. The equalization of the level population is opposed by the spin–lattice interaction which tends to restore the initial Boltzmann distribution.

It is convenient to describe the spin–lattice interaction by a relaxation time T_1 which is defined by the equation $dM_z/dt=-(M_z-M_0)/T_1$, giving the rate at which the magnetization M_z in the direction of H_0 approaches its equilibrium value M_0, if the radio-frequency field is removed. For large values of H_1, both χ' and χ'' approach zero asymptotically and the power absorption, which is given by $P=\omega V H_1=\omega\chi'' H_1^2$, approaches the limiting value given by $P_{\max}=\chi_0 H_0^2/T_1$. For water at room temperature with $H_0=2{,}000$ gauss, $T_1\simeq 5$ sec. $P_{\max}$ is of the order of 10^{-11} watts/cm^3.

More complete information about the resonance effects has been obtained from calculations of the effects of the fields on the total magnetization of the specimen (Bloch 1946, Jacobsohn and Wangsness 1948). Associated with the magnetization M there is an angular momentum A given by $M=\gamma A$ (where $\gamma=\mu/(Ih/2\pi)$ is the nuclear gyromagnetic ratio) so that, under the influence of H_0 alone, M precesses about H_0 with angular velocity $\omega=\gamma H_0$. This gyroscopic precession is influenced by the rotating couple due to H_1, and the resulting motion depends on the initial conditions and the rate of variation of the fields. By neglecting spin–spin and spin–lattice interactions, simple solutions can be obtained for some cases (Bloch 1946, Bloembergen 1948 a). If H_0 or ν are far from resonance, M is in the direction of H_0. If resonance is approached slowly, the angle of precession θ between M and H_0 increases so that M has a rotating component M_t in the transverse plane (Figure 3). Owing to the effects of spin–spin and spin–lattice interaction, the magnitude of M does not remain constant during the passage through resonance. The effects of these interactions are represented by two damping terms which are introduced into the equations. The spin–lattice interaction tends to restore the magnetization to its original value in the direction of H_0. The spin–spin interaction and field inhomogeneities tend to destroy the coherence of the phase of the spins and so reduce the transverse

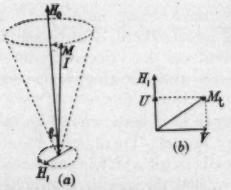


Figure 3. (*a*) Illustrating the precession of M about H_0, perturbed by the rotating field H_1. (*b*) M_t is the component of M on the transverse plane, U and V are the components of M_t.

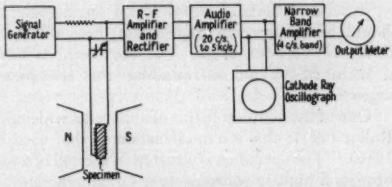


Figure 4. Experimental arrangement for observing nuclear resonance.

magnetization. The rate of decay of the transverse magnetization may be specified by a relaxation time T_2 which is defined in a way similar to T_1. The damping of the transverse magnetization is related to the width of the resonance, in fact, $\Delta\nu=1/2\pi T_2$.

Solutions of the equations, including the interaction effects, have been obtained for two cases of special interest: (i) when the amplitude of H_1 is constant and H_0 is varied slowly through resonance so that equilibrium is at all times established, (ii) when the amplitude of H_1 is constant and H_0 is varied so rapidly through resonance that the magnitude of M remains nearly constant.

For case (i) values of the components U and V of the transverse magnetization for different values of H_1 are shown in Figure 2. The values of U and V are for the optimum frequencies: $\nu=\nu_0$ for V, and $\nu=\nu_0-(2\pi T_2)^{-1}[1+(\gamma H_1)^2T_1T_2]^{\frac{1}{2}}$ for U. For small H_1 both U and V increase linearly with H_1 but, with increasing H_1, V passes through a maximum whereas U attains a limiting value. The maximum value attained by U or V is $\frac{1}{2}M_0(T_2/T_1)^{\frac{1}{2}}$.

For case (ii), when H_1 is large enough for the orientation of M to follow the changes of H_0 adiabatically, M is turned completely over on passage through resonance. The conditions which must be satisfied for this to take place are $\gamma H_1 \gg (1/H_1)(dH_0/dt) \gg 1/T_2$. If $\gamma H_1 \gg \Delta\nu$ a value of dH_0/dt can always be found which satisfies these conditions. On exact resonance, the transverse magnetization is nearly equal to M_0 so that quite a large voltage can be induced in a coil surrounding the specimen. As an example we may quote the signal to be expected from the protons in 1 cm³ of water due to a single rapid passage through resonance. For $H_0 = 2{,}000$ gauss, $M_0 = 7 \times 10^{-7}$ E.M.U. and this moment, rotating at the resonant frequency of about 4 Mc/s., will produce a voltage of several millivolts across the coil of a tuned circuit of $Q \simeq 100$. In order to observe the transient signal one would require a receiver of about 1 kc/s. bandwidth. The Johnson noise of the circuit would be less than a microvolt so that the signal would be well above noise level.

In practice it is usually more convenient to use the method of slow passage through resonance. Although the maximum signal is often much smaller (e.g. in water about 50 times smaller) this is offset by the fact that for slow passage it is possible to use a receiver with a much smaller bandwidth.

§ 4. EXPERIMENTAL METHODS

Several different experimental arrangements have been employed. It should be emphasized that, while these differ in detail, there is no essential difference between them. All the methods depend on the effects produced in a tuned circuit surrounding the specimen by the rotating transverse magnetization.

One of the simplest forms of apparatus which may be used is shown in Figure 4 (Rollin 1946); this is a modification of that used by Purcell, Torrey and Pound (1946). The specimen is situated in the coil of a tuned circuit which is connected, through a high impedance, to a signal generator, so that the circuit is supplied with a constant current. When the field H_0 is adjusted to the resonant value, the nuclear absorption causes a small drop in the shunt impedance of the circuit which is detected by the change of radio-frequency voltage across the circuit. The field H_0 is modulated with a small amplitude, by means of auxiliary coils, at an audio frequency so that, when H_0 is correctly adjusted, the radio frequency is modulated by an amount proportional to the slope of the resonance curve. After rectification, the resulting audio signal is passed through a narrow-band amplifier, the output of which is indicated by a meter. When investigating the width and shape of the resonance it is often convenient to use a larger field modulation so that the field sweeps right through the resonant region, and the absorption line can be displayed on an oscillograph.

If the circuit is tuned exactly to the signal generator the impedance is influenced, to a first approximation, only by the nuclear absorption and not by the dispersion. The signal voltage is therefore proportional to the product of χ'' and the radio-frequency voltage across the circuit, i.e. it is proportional to the component V of the transverse magnetization. If, however, the circuit is detuned from exact resonance, its impedance is altered by changes of χ' and signals can be obtained which depend mainly on the component U.

The arrangement shown in Figure 4 is not very suitable for use in experiments where high radio-frequency fields are required. In the method used by Purcell

et al. (see Bloembergen, Purcell and Pound 1948) the circuit containing the specimen is placed in one arm of a radio-frequency bridge so that only a small unbalanced signal is fed to the amplifier. By adjusting the phase balance of the bridge either absorption or dispersion can be observed. In order to obtain a narrow audio bandwidth with minimum noise, the audio signal is mixed with a voltage derived from the generator which supplies the field modulation, and the resultant signal indicated by a long period meter.

In the method of nuclear induction devised by Bloch, Hansen and Packard (1946) a separate transmitter and receiver circuit are used. In order to cut down the direct pick-up from the transmitter, the two coils are placed at right angles. The signal induced in the receiver coil by the rotating nuclear magnetization is mixed with the voltage picked up from the transmitter. Signals proportional to either U or V can be obtained by use of a device which alters the phase of the mixing voltage.

An automatic recording spectrometer for measurements on nuclear resonance has been used by Pound (1947 a). The specimen is placed in the coil of an oscillator and absorption detected by the change in amplitude of oscillation. The frequency of the oscillator is varied very slowly by a motor driven device and the output indicated on a recorder. As the passage through resonance is slow, a very narrow bandwidth may be used with consequent reduction of the noise. In experiments by other workers (e.g. Roberts 1947 a) a super-regenerative oscillating receiver has been employed.

§ 5. MEASUREMENT OF THE SPIN–LATTICE RELAXATION TIME

For slow passage through resonance, the value of H_1 for which V is a maximum is given by

$$H_1 = \frac{1}{\gamma(T_1 T_2)^{\frac{1}{2}}} = \frac{1}{\gamma}\left(\frac{2\pi\Delta\nu}{T_1}\right)^{\frac{1}{2}}$$

so that, by measuring the line width and determining this value of H_1, the spin–lattice relaxation time can be calculated.

Another method, which can profitably be used if T_1 is longer than a few seconds, depends on direct observation of the equilibrium time. The specimen is first subjected to a strong radio-frequency field so that the population of the levels is nearly equalized. The magnitude of H_1 is then reduced and the time required to attain the new equilibrium population can be found directly from the rate of increase of the signal.

Other very ingenious ways of measuring T_1 have been used by Bloch, Hansen and Packard (1946) depending on the rate of change of phase of signals obtained with a sinusoidal field modulation when the mean value of the field is suddenly changed.

§ 6. LINE WIDTH

One of the earliest observations was the existence of a remarkable difference between the width of the resonance in the solid and liquid state. In general, there are several contributions to the line width, but, in a substance in which there is no electron paramagnetism and the nuclei have no electric quadrupole interaction with internal electric fields, the major contribution is due to the mutual magnetic interaction of the spins. This interaction may be considered to influence the line width for two reasons: (i) there is a field contribution by neighbouring

nuclei so that each nucleus is in a slightly different field, (ii) the exchange of quanta between neighbouring nuclei shortens the lifetime in a state and therefore produces a broadening of the levels. The two effects are of the same order of magnitude. If we consider a simple case, such as an ionic crystal, and assume that the nuclei are nearly at rest, the static field at one nucleus due to another is approximately μ/d^3 where d is the interatomic distance. The rate of quantum interchange is $^2/\mu\hbar d^3$ so that, from the two contributions, there arises a line width $\Delta H \simeq \mu/d^3$ or $\Delta\nu \simeq \mu^2/\hbar d^3$. In solids the value of ΔH is generally found to be of the order of a few gauss, in good agreement with this calculation. In other cases, however, the line is found to be considerably narrower, for example, in water it is certainly less than a few hundredths of a gauss. The explanation of this result has been given by Bloembergen, Purcell and Pound (1948) who have pointed out that the reduced width is due to molecular motions. In the case of a liquid the molecules are undergoing the Brownian motion so that the local field acting on a nucleus is varying in a random manner. In order to calculate the width it is necessary to make a Fourier analysis of the field fluctuations and select the components near zero frequency which cause secular perturbation. The magnitude of these components depends on the correlation time of the fluctuations which, for a liquid such as water, is approximately equal to the Debye relaxation time which appears in the theory of dielectric dispersion. This theory is in quantitative agreement with the observed results and predicts a variation of width with viscosity which has been checked in the case of glycerin. The width observed in solids is also sometimes considerably less than that which would be expected if the molecules were simply undergoing lattice vibrations. These narrow lines are taken to indicate the existence of free molecular rotation. The study of nuclear resonance can provide useful information about the nature of molecular motions; for example, it has been found by Bitter *et al.* (1947a) that the transition which occurs in methane at 20° K. is due to a change from free to hindered rotation whereas the transition in ammonium chloride is not. It has also been found that there is a difference between the molecular motions in natural and synthetic rubber (Alpert 1947). Very considerable changes in line width and shape have been observed in solid ortho hydrogen between 20° and 1° K. and, at the lowest temperature, the line splits into three peaks (Rollin and Hatton 1948). This splitting is believed to be due to spin–spin interaction of the two protons in the molecule but is not yet completely explained.

For nuclei with $I > \frac{1}{2}$ the major contribution to the line width may be due to quadrupole effects (Pound 1947 b, 1948 b, c).

§7. SPIN LATTICE EQUILIBRIUM

The study of the spin–lattice relaxation time T_1 has provided an interesting field for both theoretical and experimental work. Essentially the problem is to determine the probability of the exchange of quanta between the nuclei and the thermal motions of the molecules. For exchange to take place there must be a perturbation of the nuclei which depends on the thermal motions. The simplest case to consider is again that of an ionic crystal in which there are no molecular rotations and the only perturbation is the magnetic field due to neighbouring nuclei. This magnetic field is varying on account of the relative motion of the nuclei caused by the lattice vibrations. The components of the field which are transverse to H_0 and have the Larmor frequency ν_0 are effective in inducing nuclear transitions in which energy is taken from, or absorbed by, the lattice.

The theory of relaxation by this process has been worked out by Waller (1932) and leads to a value of T_1 of the order of 10^8 seconds at room temperature and 10^{17} seconds at 1° K. These predictions are in complete disagreement with experimental results both as regards the order of magnitude and the temperature variation. The deviation is especially marked at low temperatures where values of the order of a second have been found in lithium fluoride and calcium fluoride (Rollin and Hatton 1947, 1948). The discrepancy is not due to any fault in the theory but to the fact that the relaxation time is affected by minute quantities of paramagnetic impurity or the presence of F centres; even concentrations as low as one in a million can exert a profound influence. It seems probable that the paramagnetic impurities affect the relaxation process through the fluctuating fields due to changes of electron spin orientation (Bloembergen 1948b, Rollin and Hatton 1948) which bring about equilibrium of the nuclei in the immediate neighbourhood. The remaining nuclei then reach equilibrium by conduction through the spin system.

The relaxation time in liquids has been calculated by Bloembergen, Purcell and Pound by obtaining the component of the random local field fluctuations at the Larmor frequency. The calculations are in very good agreement with experiment. The results obtained for the variation of relaxation time with viscosity in glycerin are especially interesting. At first the relaxation time decreases with increasing viscosity then, after passing through a minimum, it increases again. This is exactly what the theory predicts, the minimum being reached when the fluctuating fields at the frequency ν_0 are a maximum, which happens when the Debye time is of the order $1/2\pi\nu_0$.

In the case of liquids the effect of paramagnetic impurities can be easily verified. The first experiments were made by Bloch, Hansen and Packard (1946) and measurements have also been made by Bloembergen, Purcell and Pound (1948) who have verified the proportionality of the effect to the concentration and to the square of the effective moment of the paramagnetic ions. It has also been verified that, for $I>\frac{1}{2}$ the relaxation time is reduced by quadrupole interactions (Pound, 1947b, 1948b).

Experiments on hydrogen gas have been made by Bloembergen *et al.* (1948). For a gas the correlation time of the fluctuations is the interval between collisions. The field fluctuations are due to change of molecular orientation, which alters both the field exerted on one of the protons in the molecule by the other and also the field due to molecular rotation. In agreement with the fluctuation theory, the relaxation time becomes longer with increase of pressure, since this increases the collision frequency, which is already much greater than the Larmor frequency.

In a metal, the nuclei should be able to exchange quanta with the conduction electrons. The probability of this process has been calculated by Heitler and Teller (1936). Measurements have been made recently at 1° K. on copper and aluminium (Rollin and Hatton 1948). A value of about 1 second was obtained which is in reasonable agreement with the theory.

§8. MEASUREMENTS OF NUCLEAR MOMENTS

Paramagnetic resonance provides a useful alternative to the molecular beam method for the accurate measurement of nuclear moments.

Measurements of the magnetic moments of the light nuclei are of fundamental importance for the information which they provide about nuclear structure and the nature of nuclear forces. Several experiments have been made recently

in order to obtain a more accurate comparison of the moments of the proton, the neutron and the deuteron. The measurements are made in the same magnetic field, so that the result depends only on the comparison of frequency which can be made with very great accuracy.

In experiments made by Arnold and Roberts (1947) the method of Alvarez and Bloch (1940) was employed for the neutron determination and the proton and deuteron moments were measured by observing the resonances in a water sample. The neutron moment has been recently redetermined (Bloch, Nicodemus and Staub 1948) using a method in which the magnetic field is stabilized by using the proton resonance (Packard 1948). The proton–deuteron ratio has also been measured by Roberts (1947b) and Bloch, Levinthal and Packard (1947) using water samples and by Bitter *et al.* (1947c) using liquid hydrogen and deuterium. Very high accuracy can be achieved by using a sample containing both protons and deuterons so that, by applying two frequencies simultaneously, both resonances may be displayed on an oscillograph screen and the frequencies then adjusted until the exact resonance points are in coincidence. By using this method the ratio of the moments can be obtained to 1 part in 10^5. One of the difficulties in these measurements is to locate the point on the oscillograph pattern which corresponds to exact resonance. Unless the rate of sweep through resonance is very slow the pattern seen does not resemble an ordinary resonance curve. The maximum signal is obtained just after passing through the true resonance point and this is then followed by a number of oscillations, due to the fact that the transverse magnetization persists for some time after passage through resonance and, as it rotates with the instantaneous Larmor frequency, comes in and out of phase with the radio-frequency field. The theory of these transient phenomena has been considered by Bitter *et al.* (1947b) and by Jacobsohn and Wangsness (1948). It may be shown, that if a sinusoidal sweep is used, so that the patterns seen for the forward and reverse sweep are symmetrical, the exact resonance is located at the centre of symmetry.

Recently a measurement has been made of the moment of the radioactive isotope of hydrogen ^{3}H (Anderson and Novick 1947, Bloch, Graves, Packard and Spence 1947). The specimen used was a drop of water made from ^{3}H obtained from a chain reacting pile. It was expected that the moment would be slightly lower than that of the proton due to the fact that it is not in a pure S state (Sachs and Schwinger 1946) but, in fact, it was found to be about 7% greater. This discrepancy has been ascribed to the existence of an exchange contribution to the magnetic moment (Villars 1947). Further important information has been obtained from the measurement of the moment of ^{3}He by Anderson and Novick (1948). A sample of ^{3}He obtained from the decay of ^{3}H was mixed with an equal volume of oxygen and compressed to 20 atmospheres. The object of adding the oxygen was to accelerate the attainment of nuclear equilibrium by the influence of the fields due to the paramagnetic molecules. The moment of ^{3}He was found to be numerically greater than that of the neutron, in agreement with the prediction of Villars. From the moments of ^{3}H and ^{3}He information is obtained about the mixture of nuclear states (Sachs and Schwinger 1946, Sachs 1946) and the character of nuclear forces (Sachs 1948, Anderson 1948). Other nuclei of which the moment has been measured by the paramagnetic resonance method include: bromine (Pound 1947b), iodine, gallium and phosphorus (Pound 1948b), copper (Pound 1948a) and thallium (Poss 1947).

§ 9. RESONANCE IN SINGLE CRYSTALS

In experiments on single crystals, anisotropic effects and a splitting of the resonance have been observed in some cases. These phenomena are due, either to magnetic dipole–dipole interaction of spins, or to electric quadrupole interaction with field gradients. In calcium fluoride it is found that the width of the resonance depends on the orientation of H_0 to the crystallographic axes, due to the variation of spin–spin interaction (Purcell, Bloembergen and Pound 1946). Very interesting results have been obtained by Pake (1948) in measurements in a crystal of gypsum ($CaSO_4 . 2H_2O$). It is found that the proton resonance is split into four lines due to the spin–spin interaction of the two pairs of protons in the two water molecules. By measuring the separation of the peaks, the distance between the protons can be calculated.

The effect of quadrupole interactions in single crystals has been investigated by Pound (1948c, d). In cubic crystals there is no electric field gradient at the nucleus, so that quadrupole effects are found only in crystals with lower than cubic symmetry. In $Li_2SO_4 . H_2O$ the ^7Li resonance was found to be split into a triplet and in a corundum crystal (Al_2O_3) the ^{27}Al line was found to have the five components characteristic of a spin of 5/2.

§ 10. NUCLEAR PARAMAGNETISM AND LOW TEMPERATURES

There are several interesting possibilities in connection with the application of nuclear paramagnetism in low temperature research. Investigations have been made recently to explore the possibility of using the nuclear susceptibility of a substance for measurement of temperature in the region below 0·1° K. The use of electron paramagnetism for temperature measurement is limited by the deviations from Curie's law due to the effects of the crystalline electric fields and magnetic and exchange interactions of the spins. For the nuclear susceptibility of a substance in which there are no quadrupole interactions, deviations from Curie's law will arise only on account of the magnetic spin–spin interaction, which will not become apparent until temperatures of the order of 10^{-5} °K. are attained. In order that the nuclear susceptibility should be a true indication of the temperature of the substance it is essential for the spin lattice equilibrium time to be reasonably short. From recent measurements (Rollin 1948) it is apparent that this condition is satisfied, at least down to a temperature of 0·2° K.

The use of adiabatic nuclear demagnetization for obtaining very low temperatures has been suggested by Gorter (1934) and by Kurti and Simon (1935). It has been estimated by Simon (1940) that the lowest useful temperature which can be attained using electron paramagnetism is of the order of 10^{-4} °K. The temperature which can be reached is determined by the position of the specific heat anomaly due to electron spin–spin interaction. Even if temperatures lower than the anomaly were reached, it would not be possible to make any useful measurements in this region, as the specific heat of the substance would be so low that the temperature would rise rapidly, due to the unavoidable heat inflow. The temperature at which the anomaly occurs can be reduced by using a substance in which the concentration of paramagnetic ions is very low, but a limit is set to the dilution by the reduction of magnetic entropy of the substance and the corresponding reduction of specific heat in the region of the anomaly. This difficulty can be avoided by the use of nuclear paramagnetism, because the interaction between the nuclei is very small but the entropy change, per unit volume,

associated with orientation of the spins is very high. In order that the nuclear demagnetization should be successful it is necessary to remove an appreciable amount of entropy on magnetization, that is, to produce an appreciable degree of nuclear alignment. This can only be done by the use of very high magnetic fields at very low temperatures. If a field of 50 kilogauss were available the magnetization should be carried out at a temperature of about a hundredth of a degree, so that it would be necessary to use an initial stage of electron demagnetization. Further progress in this direction is therefore dependent on the successful development of a two-stage demagnetization process. As regards the choice of a suitable substance, the requirements are that it should have a high volume susceptibility and a reasonably short relaxation time. It seems that these conditions are most likely to be fulfilled by metals such as aluminium, copper or indium. Apart from the possibility of reaching very low temperatures, the achievement of nuclear alignment would afford an opportunity for studying nuclear properties such as the angular distribution of particles or radiation emitted during radioactive disintegration. Theoretical considerations (Spiers 1948) indicate that for $I > \frac{1}{2}$ appreciable anisotropic effects should be obtained.

It has recently been suggested (Gorter 1948) that it should be possible to align the nuclei of certain paramagnetic ions by making use of the large internal field ($\sim 10^6$ gauss) acting on the nucleus due to the electronic moment. At a temperature of 10^{-2} °K. almost complete alignment would be obtained with an external field of only a few hundred gauss.

REFERENCES

ALPERT, N. L., 1947, *Phys. Rev.*, **72,** 637.
ALVAREZ, L. W., and BLOCH, F., 1940, *Phys. Rev.*, **57,** 352.
ANDERSON, H. L., 1948, *Phys. Rev.*, **73,** 919.
ANDERSON, H. L., and NOVICK, A., 1947, *Phys. Rev.*, **71,** 372; 1948, *Ibid.*, **73,** 919.
ARNOLD, W. R., and ROBERTS, A., 1947, *Phys. Rev.*, **71,** 878.
BITTER, F., ALPERT, N. L., NAGLE, D. E., and POSS, H. L., 1947c, *Phys. Rev.*, **72,** 1271.
BITTER, F., ALPERT, N. L., POSS, H. L., LEHR, C. G., and LIN, S. T., 1947a, *Phys. Rev.*, **71,** 738; 1947b, *M.I.T. Quarterly Progress Report*, 15th July 1947.
BLOCH, F., 1946, *Phys. Rev.*, **70,** 460.
BLOCH, F., HANSEN, W. W., and PACKARD, M. E., 1946, *Phys. Rev.*, **70,** 474.
BLOCH, F., GRAVES, A. C., PACKARD, M. E., and SPENCE, R. W., 1947, *Phys. Rev.*, **71,** 373.
BLOCH, F., LEVINTHAL, E. C., and PACKARD, M. E., 1947, *Phys. Rev.*, **72,** 1125.
BLOCH, F., NICODEMUS, D., and STAUB, H. H., 1948, *Phys. Rev.*, **74,** 1025.
BLOEMBERGEN, N., 1948a, *Nuclear Magnetic Relaxation* (The Hague : Martinus Nijhoff); 1948b, *Proc. Phys. Soc.*, **61,** 575 (abstract only).
BLOEMBERGEN, N., PURCELL, E. M., and POUND, R. V., 1948, *Phys. Rev.*, **73,** 679.
FRENKEL, J., 1945, *J. Phys. U.S.S.R.*, **9,** 299.
GORTER, C. J., 1934, see DEBYE, P., *Phys. Z.*, **35,** 923.
GORTER, C. J., 1936, *Physica*, **3,** 995 ; 1948, *Ibid.*, **14,** 504.
GORTER, C. J., and BROER, L. J. F., 1942, *Physica*, **9,** 591.
GORTER, C. J., and KRONIG, R. DE L., 1936, *Physica*, **3,** 1009.
HEITLER, W., and TELLER, E., 1936, *Proc. Roy. Soc.* A, **149,** 152.
JACOBSOHN, B. A., and WANGSNESS, R. K., 1948, *Phys. Rev.*, **73,** 942.
KURTI, N., and SIMON, F., 1935, *Proc. Roy. Soc.* A, **155,** 629.
LASAREW, B. G., and SCHUBNIKOW, L. W., 1937, *Phys. Z. Sowjet*, **11,** 445.
PACKARD, M. E., 1948, *Rev. Sci. Instrum.*, **19,** 435.
PAKE, G. E., 1948, *J. Chem. Phys.*, **16,** 327.
POSS, H. L., 1947, *Phys. Rev.*, **72,** 637.

POUND, R. V., 1947a, *Phys. Rev.*, **72,** 527; 1947b, *Ibid.*, **72,** 1,273; 1948a, *Ibid.*, **73,** 523; 1948b, *Ibid.*, **73,** 1,112; 1948c, *Ibid.*, **73,** 1,247; 1948d, *Proc. Phys. Soc.*, **61,** 576 (abstract only).
PURCELL, E. M., BLOEMBERGEN, N., and POUND, R. V., 1946, *Phys. Rev.*, **70,** 988.
PURCELL, E. M., TORREY, H. C., and POUND, R. V., 1946, *Phys. Rev.*, **69,** 37.
ROBERTS, A., 1947a, *Phys. Rev.*, **72,** 182; 1947b, *Ibid.*, **72,** 979.
ROLLIN, B. V., 1946, *Nature, Lond.*, **158,** 669; 1948, *Proc. Phys. Soc.*, **61,** 575 (abstract only).
ROLLIN, B. V., and HATTON, J., 1947, *Nature, Lond.*, **159,** 201; 1948, *Phys. Rev.*, **74,** 346.
ROLLIN, B. V., HATTON, J., COOKE, A. H., and BENZIE, R. J., 1947, *Nature, Lond.*, **160,** 436.
SACHS, R. G., 1946, *Phys. Rev.*, **69,** 611; 1948, *Ibid.*, **73,** 1,222.
SACHS, R. G., and SCHWINGER, J., 1946, *Phys. Rev.*, **70,** 41.
SIMON, F., 1940, *Le Magnetisme III* (Paris: Institut International de Coopération Intellectuelle).
SPIERS, J. A., 1948, *Nature, Lond.*, **161,** 807.
VILLARS, F., 1947, *Phys. Rev.*, **72,** 256.
WALLER, I., 1932, *Z. Phys.*, **79,** 370.
ZAVOISKY, E., 1945, *J. Phys. U.S.S.R.*, **9,** 211, 245, 447.

PHOSPHORS AND PHOSPHORESCENCE

By G. F. J. GARLICK
Department of Physics, University of Birmingham

CONTENTS

PAGE

§1. Introduction 35
The theoretical model for photoconducting phosphors 36
§2. Absorption and emission characteristics of phosphors 37
(i) Absorption spectra of sulphide and silicate phosphors 37
(ii) Excitation spectra of phosphors 38
(iii) Emission spectra of phosphors 39
(iv) The luminescence efficiency of phosphors 40
§3. Phosphorescence and thermoluminescence 41
(i) The thermoluminescence experiment 41
(ii) Phosphorence due to electron traps 43
(iii) Retrapping of electrons during phosphorescence and thermoluminescence 44
(iv) Phosphorescence of short duration due to electron traps 46
(v) The electrical characteristics of electron traps 46
(vi) The nature of the electron traps in phosphors 46
§4. Luminescence and phosphor constitution and preparation 47
(i) Zinc and zinc–cadmium sulphide phosphors 47
(ii) The alkaline earth sulphide phosphors 48
(iii) Silicates and other oxygen-containing phosphors 49
(*a*) Silicate phosphors 49
(*b*) Tungstate phosphors 51
(*c*) Zinc oxide and aluminium oxide phosphors 51
(iv) Tetravalent manganese as a luminescence activator 51
§5. Infra-red sensitive phosphors 51
Theoretical studies of infra-red stimulation 53
§6. Conclusion 53

ABSTRACT. The report reviews recent investigations of the luminescence characteristics of impurity activated phosphors and the basic processes involved. The development of the electron energy band model of crystalline phosphors and experimental support for it is discussed. Phosphorescence of long duration has been shown to be due to the trapping of electrons in thermally metastable levels in phosphors : correlation of phosphorescence and thermoluminescence due to such electron traps has been achieved. The nature of traps is only known for a few types of phosphor. There have been important advances in the knowledge of the structure of luminescence emission centres in sulphide and silicate phosphors : in zinc sulphide phosphors with copper or silver impurity the centres are thought to be based on molecular halide groups formed by the impurity and the halogen which has to be included to obtain luminescence at all. A section is devoted to the relation of the emission spectra of manganese-activated silicate phosphors to the state of coordination of the manganese and the action of certain additional impurities, such as arsenic and tin, in promoting phosphorescence of long duration in these materials. Recent studies of oxide and tungstate phosphors are mentioned.

A brief summary of the recent development of phosphors sensitive to the stimulating action of infra-red radiation is given. It is found that a secondary activator is necessary to obtain marked stimulability. Although experimental data on the basic processes are sparse, it appears that infra-red light results in the ejection of trapped electrons. However, there seems to be no simple correlation between the process of optical ejection of trapped electrons and the thermal ejection which takes place during thermoluminescence.

§1. INTRODUCTION

DURING the last decade a large increase in the number of studies of luminescence has occurred and it is not possible to provide a complete review of all aspects of the subject. The present report is therefore restricted to a survey of investigations of impurity-activated phosphors of crystalline structure which are synthesized by heat treatment of their basic constituents. In particular we shall consider sulphide and silicate phosphors. These solids possess the most interesting luminescence features, such as phosphorescence, thermoluminescence and such related characteristics as photoconductivity and dielectric changes accompanying luminescence. In the remainder of this introduction we shall provide a brief review of the state of knowledge prior to 1939 when the last report appeared (Randall and Wilkins 1939) and when the last British symposium on luminescence had just terminated (Faraday Society 1939).

We consider first the conditions for luminescence to occur in solids. It was recognized that the first requisite for luminescence in solids is the existence of certain localities in the crystal where radiative processes can take place following absorption of energy; the structure of these *luminescence centres* provides the shielding from lattice vibrations which would otherwise remove the energy by thermal dissipation (Peierls 1932). In some particular cases these centres could be specified, as shown by Randall (1939). It was known that in the manganese halides the Mn^{2+} ions formed the emission centres while in some salts, such as uranyl salts, a coordination group formed the emission centre. In the zinc and alkaline earth sulphides the nature of the luminescence centres was much less well defined although it was clear that they were due to the introduction of the impurity activators or to non-stoichiometric composition of the matrix lattice. It was thought more likely that the impurity centres occupied interstitial positions in the lattices (Riehl 1939a). With respect to the phenomena of phosphorescence and photoconductivity, a clear distinction was made between photoconducting and non-photoconducting phosphors. When the luminescence process is confined entirely to the centres in which emission occurs then the mechanism of decay is easy to understand and no photoconductivity is expected. In this case the simplest decay arises from the forbidden nature of the electron transitions giving the emission. It is exponential in form and its rate is independent of the temperature and of the excitation conditions (Garlick and Wilkins 1945). Metastable states of the luminescence centre may give rise to phosphorescence of longer duration, as in the case of thallium centres in potassium chloride containing the heavy metal as impurity. The temperature-dependent, exponential decay of phosphorescence is easily related to the metastable state of the centre having a single activation energy. Most of the studies of such phosphors were due to Bünger and Flechsig (1931) and the establishment of a model for the alkali halide phosphors to Seitz (1938).

The knowledge of the phosphorescence processes in photoconducting phosphors was not as precise as that just described. It was assumed from the conductivity changes that electrons were removed from luminescence centres by excitation and that the phosphorescence observed was due to their relatively slow rate of recombination with the empty luminescence centres. The existence of trapping states in the crystal lattice which could also delay the return of electrons to centres was assumed to account for phosphorescence of very long duration.

However, no precise knowledge of these states was available and the phosphorescence was treated as analogous in mechanism to a simple bimolecular reaction. Hardly any of the experimental measurements of phosphorescence decay justified this simple picture. In most cases of hyperbolic decay of phosphorescence the simplest empirical equation was of the following form:

$$I = I_0/(1+at)^n \qquad \ldots\ldots (1)$$

where I is the phosphorescence intensity at decay time t, I_0 its initial value and a and n constants. The constant n was usually less than the value two expected from bimolecular theory and varied between one and two for different phosphors. The decay of the conductivity, to be expected during phosphorescence, was treated in a simple manner by Reimann (1937) but no reliable experimental measurements were available. The dielectric changes during luminescence in these phosphors were assumed to be due to the conduction electrons in the phosphor (Gisolf 1939, de Groot 1946). The first application of a comprehensive

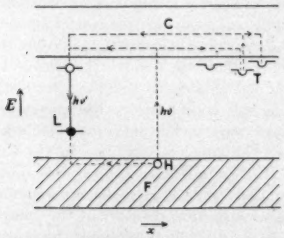


Figure 1. Schematic electron energy diagram for an impurity activated crystalline phosphor.
C. Conduction energy band.
F. Filled energy band.
H. Positive hole.
L. Luminescence centre.
T. Electron traps.
$h\nu$. Absorbed quantum.
$h\nu'$. Emitted quantum.

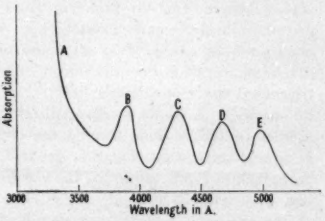


Figure 2. Absorption spectrum of a ZnS–Mn 2% phosphor (Kröger).
A. ZnS matrix.
B, C, D and E. Bands due to Mn^{++} centres.

theoretical model to explain the general behaviour of such phosphors was made in 1939 by Johnson (1939) and independently by Riehl (1939b). Both made use of the collective electron model for crystalline insulators originally developed by Bloch (1928). It had been successfully applied to the alkali halides and its general usefulness shown by Mott and Gurney (1940). We consider it very briefly below as it forms the basis of most subsequent treatments of luminescence and related phenomena in phosphors.

The theoretical model for photoconducting phosphors.

The energy states of the outermost electrons of the phosphor matrix atoms fall into allowed bands separated by regions of forbidden energy states as shown in Figure 1. The highest occupied band is normally completely saturated in insulators while the allowed band above is normally empty. This band is known as the conduction band since electrons raised into it by excitation can contribute

to a current through the crystal. Impurities and lattice defects give rise to localized levels as shown in the Figure. These may lie in the forbidden bands of the matrix. Absorption in the matrix lattice excites electrons from the filled band into the conduction band. Excitation states may also be reached which represent electrons still bound to their positive holes (Frenkel 1931). We shall not consider these. Absorption in the impurity centres may either raise electrons into the conduction band or into excited states of the centres. When positive holes occur in the filled band they may capture electrons from unexcited luminescence centres and so empty the latter. The trapping states shown as levels just below the conduction band, and arising from lattice defects and other disturbances in the crystal, will only show absorption characteristics when they are filled by electrons, that is, during or after excitation. Their absorption will be in spectral regions of relatively long wavelength. We see from this model that luminescence will occur whether excitation involves the matrix lattice atoms or only the luminescence centres since in both cases the luminescence centres are emptied of electrons which later recombine with them. Positive hole motion is, in the first case, a constituent part of the luminescence mechanism. It is obvious that this model for phosphors can be tested by measuring the absorption characteristics of particular types of phosphors. We shall review such measurements in § 2 (i).

§ 2. ABSORPTION AND EMISSION CHARACTERISTICS OF PHOSPHORS

(i) *Absorption Spectra of Sulphide and Silicate Phosphors*

Most of the precise measurements of the absorption spectra of these phosphors have been due to Kröger and Gisolf and their associates in Eindhoven. Kröger (1940a) has provided a fairly complete account of these studies and has discussed their relation to the above theoretical model. We consider first the zinc sulphide phosphor systems. For pure zinc sulphide the lattice absorption begins at a wavelength of about 3400 A., as shown by Kröger (1939a) and others (see Brasefield 1940), and the absorption has coefficients of the order of $10^4\ \mathrm{mm}^{-1}$. At longer wavelengths this absorption has a decreasing tail with absorption coefficients of the order of $1\ \mathrm{mm}^{-1}$ and is suggestive of absorption by lattice irregularities and by interstitial atoms. A discussion of this part of the absorption spectrum occurred in the Faraday Society (1939) Symposium. The impurity activator also gives rise to absorption in this region. In particular we may mention manganese whose specific absorption is shown in Figure 2 together with that of the matrix lattice of zinc sulphide. Since manganese sulphide forms mixed crystals with zinc sulphide the substitution of manganese for zinc in the zinc sulphide lattice causes a shift of the lattice absorption edge to longer wavelengths. The absorption edge for manganese sulphide is at 3650 A. A similar shift in the absorption is found for mixed crystals of zinc and cadmium sulphides (Kröger 1940b). The absorption edge for cadmium sulphide is at 5170 A. The absorption spectra characteristic of transitions within the manganese ions are relatively well defined, as shown by the long wavelength bands of Figure 2. When absorption occurs in these bands the luminescence process is entirely confined to the manganese centres and the emission characteristics, such as the phosphorescence, are simple in nature. The phosphorescence of these phosphors for excitation in different spectral regions has been measured by Garlick and Wilkins (1945).

When excitation involves the tail of the fundamental absorption or is within the fundamental absorption band then phosphorescence due to electron trapping is observed. This would appear to give direct support to the above theoretical model since only electrons removed from the centres should be capable of being trapped.

Similar results are obtained for zinc silicate phosphors containing manganese as activator since zinc silicate and manganese orthosilicate form mixed crystals (Kröger 1939b). The absorption edge of zinc silicate is at 2250 A. while that for manganese orthosilicate is at 3000 A. The zinc silicates do not usually show long afterglow of any intensity but some specimens are known which possess such phosphorescence. They contain a small amount of arsenic. Schulman (1946) has attempted an explanation of the action of this additional impurity. We shall discuss this later. Returning to absorption measurements on zinc sulphides, it is found that other impurity activators such as copper and silver do not give rise to specific absorption bands like those of the Mn^{2+} ions. These impurities are probably interstitial and in themselves possess no features favourable to luminescence but only when in suitable lattice environments. Figure 3 (de Groot, Gisolf and Kröger 1941) shows measurements of the absorption spectra of copper activated zinc sulphide phosphors which have a definite peak due to the activator; measurements by other workers have not shown this to be a general feature. All that is usually evident is some structureless change in the tail of the matrix absorption. This difference between the spectra of manganese and other impurity-activated zinc sulphides has some significance in determining the nature of the emission centres in the latter cases and we shall refer to it later.

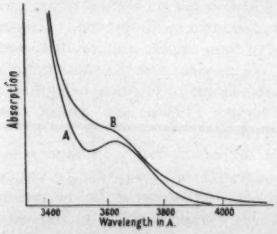


Figure 3. Absorption spectra of two different specimens of ZnS–Cu phosphors (Gisolf, de Groot and Kröger).

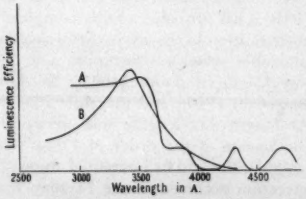


Figure 4. Excitation spectra of zinc sulphide phosphors. A. ZnS–Mn (1%) (Kröger). B. ZnS–Zn (Garlick and Gibson).

(ii) *Excitation Spectra of Phosphors*

Although we would expect the excitation spectra of phosphors to be closely related to their absorption spectra there are significant differences between them. In Figure 4 are shown the excitation spectra of a pure zinc sulphide at a low temperature (Garlick and Gibson 1948a) and also the excitation spectrum of a manganese-activated zinc sulphide at room temperature (1% Mn) (Kröger 1940a). It is characteristic of copper, silver and self-activated zinc sulphide phosphors that the excitation efficiency falls when the exciting wavelength becomes less than that of the matrix absorption edge. This effect is much less marked in manganese-activated phosphors and is often absent. It would appear that although the absorption coefficient increases to large values at shorter wavelengths

it is not the cause of the above differences. It seems probable that in copper, silver and self-activated phosphors the excitation of electrons into the conduction levels and their high density in these levels makes their non-radiative return to ground states a more probable process than luminescence. In these phosphors the maximum excitation efficiency always occurs for wavelengths just larger than that of the absorption edge of the matrix lattice.

(iii) *Emission Spectra of Phosphors*

The emission spectra of sulphide and silicate phosphors are usually broad and structureless. However, in some cases their measurement has furnished valuable information. It is known, from the studies of Kröger (1940a) and his associates, that the emission spectrum of the Mn^{2+} ion in these phosphors consists of one or more bands. For example, zinc silicate containing about 1% of manganese has a green emission due to a band with a maximum at 5300 A. As the manganese content is increased the emission at low temperatures becomes yellow and then red, due to the appearance of a new spectral band with a maximum at 6100 A. This band is characteristic of the manganese metasilicate $MnSiO_3$ at low temperatures and is also produced at room temperatures in zinc silicate phosphors having normal manganese content (1%) when these contain excess silica. Such yellow and red emitting zinc silicates have been recently studied by Kröger (1940a), by Fonda (1940) and by Froelich (1943). The yellow and red emissions also appear when conventional willemite phosphors are cooled rapidly after preparation. These phosphors have the same crystal structures as normal

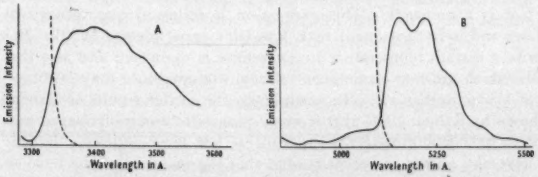


Figure 5. Edge emission spectra of zinc sulphide and cadmium sulphide phosphors at 90°K ; 2537 A° excitation (Kröger).
A. ZnS. B. CdS. Broken curve shows position of matrix absorption edge for each phosphor.

willemites (Kröger 1940a) but the identity of the red emission band with that of $MnSiO_3$ suggests that the presence of excess silica causes localized formation of $MnSiO_3$ in the lattice. In zinc–cadmium sulphide changes in the lattice constitution do not affect the position of the manganese emission band, but for copper, silver and self-activated phosphors there is a shift to the red with increase in the cadmium sulphide content (Rothschild 1937). This suggests that in the latter phosphors the luminescence centres are not formed by the impurity ions alone but by them in conjunction with the neighbouring lattice ions. We shall refer to this in a later section (§4(i)). It is important here to include a brief description of the so called " edge emission " spectra of zinc and cadmium sulphide phosphors. Kröger (1940a, b) has found that when self-activated zinc, zinc–cadmium or cadmium sulphide phosphors are excited at low temperatures by radiation of wavelength shorter than that of their matrix absorption edge, emission

occurs in a band adjacent to the absorption edge at the long wavelength side. Thus the emission is in the long wavelength ultra-violet region for zinc sulphide and in the green for cadmium sulphide. These edge emission bands are shown in Figure 5. They have a vibrational band structure, the frequency differences between adjacent maxima agreeing approximately with the 'Reststrahlen' frequencies for the respective lattices. Since they are only excited by absorption in the crystal lattice the transitions involved can only be available to electrons in the conduction band. The emission disappears at room temperature but the transitions are probably still available for conduction electrons, though now of a non-radiative character. Their presence may account for the fall in efficiency of the normal luminescence emission when the excitation is of shorter wavelength than that of the absorption edge (Garlick and Gibson 1948a).

(iv) *The Luminescence Efficiency of Phosphors*

It was assumed by Mott and Gurney (1939) that the efficiency of luminescence of a phosphor depended on the potential energy configurations of the emission centres in the normal and excited states. Using the theoretical model for these configurations developed by von Hippel (1936) and by Seitz (1939) they derived the following expression for the efficiency η:

$$\eta = A/(A+B) = 1/\{1 + c \exp(-W/\boldsymbol{k}T)\} \quad \ldots\ldots (2)$$

where A is the probability of radiative transition from excited to normal state and B is the probability of non-radiative transition between the two states: B involves the Boltzmann function $\exp(-W/\boldsymbol{k}T)$ where W is an activation energy; c is a constant. This expression is followed qualitatively by most phosphors and is in agreement with Randall's experiments (1939). It implies that above a certain temperature luminescence is quenched and also that many materials which are non-luminescent at room temperature may become luminescent at low temperatures. In addition to the earlier results of Randall it has been shown by Kröger (1947a) that many tungstates and molybdates, previously thought to be non-luminescent, can emit at low temperatures. Kröger *et al.* (1948) have also shown in recent studies that the presence of the term B in the above expression will influence the rate of decay of luminescence at higher temperatures. The decay, determined by the forbidden transition of probability A at low enough temperatures, will have a rate given by $(A+B)$ at higher temperatures. This is found to be so in experiment, that is, at temperatures where the efficiency decreases, the decay rate increases (Garlick and Wilkins 1945). At low temperatures the efficiency may vary slowly with temperature having a positive or negative temperature coefficient according to the wavelength of the exciting radiation. This slow variation and its sign have been explained in terms of the excitation spectrum of the particular phosphor and the shift of this spectrum due to thermal expansion of the matrix lattice (Garlick and Gibson 1948a).

It has been found that in photoconducting phosphors the efficiency in the region where the non-radiative transitions, that is the term B, become significant is also dependent on the intensity of the exciting radiation. This means that in this temperature region the fluorescence is not proportional to the excitation intensity. Such effects were first observed by Riehl (1939c) and by Gisolf and Kröger (1939). Theoretical explanations have been put forward by Schön

(1942) and more recently by Klasens (1946) and by Klasens and Wise (1948). Other experimental studies have been made by Urbach, Urbach and Schwarz (1947) and by Garlick and Gibson (unpublished). A useful summary of these results is given by a modified empirical form of equation (2) as follows:

$$\eta = 1/\{1 + b \exp(-W/\boldsymbol{k}T)/J^{n-1}\} \quad \ldots\ldots(3)$$

where J is the excitation intensity and n is a constant greater than unity. At temperatures where the exponential term is much greater than unity the fluorescence F is related to the excitation intensity J by the simple formula

$$F \propto J^n. \quad \ldots\ldots(4)$$

It has been found in the author's experiments that n varies between one and two according to the intensity and wavelength of the exciting light. No satisfactory theoretical explanation of the above behaviour of photoconducting phosphors has yet been given. Schön (1942) and later Klasens and Wise (1948), have assumed processes to occur in which electrons migrate through the conduction band and positive holes through the filled band, so that some emitting centres are filled non-radiatively by electrons raised from the filled band. The separation of the luminescence centre ground state from the top of the filled band determines the activation energy W. This theory can only give values of the constant n equal to 1·5 unless certain doubtful assumptions are made. Experiments made to measure the thermoelectric power of phosphors, which become semiconductors in this temperature region, show that there is no appreciable positive hole motion in the phosphor (unpublished studies of Garlick and Gibson). The evidence is based on the sign of the measured thermoelectric potentials, known semiconductors being used as standards. At present insufficient data on the above effects exist to establish the precise mechanisms involved.

§ 3. PHOSPHORESCENCE AND THERMOLUMINESCENCE

The quantitative study of the relations between phosphorescence, thermoluminescence and the electron trapping states in the theory proposed by Johnson (1939) and others began with the experiments of Randall and Wilkins (1945). As shown by them, phosphorescence and thermoluminescence are both due to the same physical process: in thermoluminescence the temperature is varied whereas in phosphorescence it remains fixed. Randall and Wilkins have shown that when phosphors are excited at low temperatures and then warmed in the dark at a uniform rate the emission varies in a characteristic way with temperature. This variation represents the electron trap distribution in the phosphor, that is, the number of electron traps of different activation energies, the latter being referred to as the trap depths. The form of the decay of phosphorescence with time has been shown to be determined by this trap distribution, and also by the extent to which traps of different depths are filled (Randall and Wilkins 1945, Garlick and Gibson 1948b). Thus the trap distributions of phosphors may be used to construct the phosphorescence decay curves at different temperatures and these may then be compared with those obtained from experiment. We shall describe first the thermoluminescence experiment and its associated theory.

(i) *The Thermoluminescence Experiment*

To carry out the thermoluminescence experiment the phosphor is mounted in a thin layer on a rhodium plated copper surface which forms part of the inner

wall of a dewar vessel of german silver. The temperature is lowered by cooling the copper surface with liquid air introduced into the inner part of the dewar vessel which has an electric heater at the rear of the copper surface. A window of glass or quartz in the outer wall enables the phosphor to be excited and the luminescence to be measured. In most experiments carried out by Randall and Wilkins (1945), and subsequently by other workers, the warming rate is chosen to be about 2·5° c/sec. which is convenient in experiment. Figure 6 gives the thermoluminescence characteristics of a long afterglow, copper-activated zinc sulphide phosphor (curve A) and also those of a phosphor containing traps of one depth only (curve B which we shall discuss below). By comparison of the two curves it is evident that the zinc sulphide has a wide distribution of electron traps. This is generally the case for phosphors of many different types and those having traps of a single depth are rare exceptions to this rule. We now consider the theory of the thermoluminescence experiment.

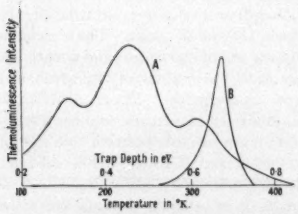


Figure 6. Thermoluminescence characteristics of a long afterglow ZnS–Cu phosphor (curve A). Curve B is the theoretical curve for a phosphor having traps of one depth only.
(E=0·67ev; S=2·9.10⁹ sec⁻¹.)

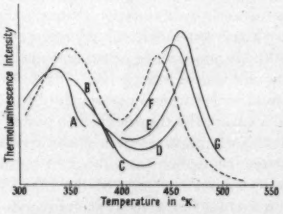


Figure 7. Correlation between the thermoluminescence curve of a SrS–Bi phosphor and curves calculated from phosphorescence at different fixed temperatures (Wilkins).
Decay at following temperatures:
A. 293° K. B. 333° K. C. 352° K.
D. 371° K. E. 390° K. F. 408° K.
G. 443° K.
------Thermoluminescence curve.

When an electron is trapped in a trap of depth or activation energy E, then it has a probability p per unit time of escaping at temperature T which is given by

$$p = s \exp(E/\boldsymbol{k}T), \qquad \ldots\ldots(5)$$

where s is a constant of the order of $10^{8\pm1}\,\text{sec}^{-1}$ for most sulphide and silicate phosphors and for many other types of phosphors. If it is assumed that the process of return of trapped electrons to empty luminescence centres does not involve retrapping then the intensity of luminescence is given by the rate of escape of electrons from traps. For only one depth of trap the intensity I when n electrons are trapped at temperature T is given by the following equations:

$$I = -dn/dt = ns \exp(-E/\boldsymbol{k}T). \qquad \ldots\ldots(6)$$

To solve these equations for the case of thermoluminescence we put $dT = \beta\, dt$ where β is the rate of warming. Then we have

$$-dn/n = (s/\beta) \exp(-E/\boldsymbol{k}T)\, dT \qquad \ldots\ldots(7)$$

which gives on integration

$$\ln n/n_0 = -\int_0^T (s/\beta)\exp(-E/\boldsymbol{k}T)\,dT \qquad \ldots\ldots(8)$$

and

$$n = n_0 \exp\left\{-\int_0^T (s/\beta)\exp(-E/\boldsymbol{k}T)\,dT\right\}. \qquad \ldots\ldots(9)$$

Thus the thermoluminescence intensity is given by:

$$I = n_0 s \exp(-E/\boldsymbol{k}T)\exp\left\{-\int_0^T (s/\beta)\exp(-E/\boldsymbol{k}T)\,dT\right\}. \qquad \ldots\ldots(10)$$

The form of this expression is given by curve B of Figure 6 for the case when s is $2{\cdot}9\times10^9\,\mathrm{sec}^{-1}$, β is $2{\cdot}5°$/sec. and E is 0·67 ev. The temperature at which the curve has a maximum does not vary rapidly with the warming rate and is found to be approximately proportional to the trap depth. The trap depth scale has been inserted in Figure 6 together with the temperature of emission. These theoretical derivations were successfully applied by Randall and Wilkins to the thermoluminescence of thallium-activated potassium chloride. Their values of E and s are as given for curve B of Figure 6 and agree with those obtained for this phosphor by phosphorescence measurements due to Bünger and Flechsig (1931). The assumption that retrapping of electrons does not occur is justified for this phosphor since the electron traps are within the luminescence centres and the phosphor shows no photoconductivity when excited. The metastable states giving rise to the trapping effects have been associated with centres formed by pairs of adjacent thallium ions (Seitz 1938). In addition to the trapping states of depth 0·67 ev. much shallower ones have been found by Randall and Wilkins (1939) which are inoperative at room temperature but modify the phosphor characteristics when it is excited at low temperatures.

(ii) *Phosphorescence due to Electron Traps*

If we neglect retrapping, the phosphorescence is given by equations (6), T being fixed. When only one depth of trap is present solution of these equation gives

$$I = n_0 s \exp(-E/\boldsymbol{k}T)\exp\{-st\exp(-E/\boldsymbol{k}T)\} \qquad \ldots\ldots(11)$$

$$= I_0 \exp\{-st\exp(-E/\boldsymbol{k}T)\} \qquad \ldots\ldots(12)$$

where I_0 is the initial intensity at zero time. This equation was first proposed by Bünger and Flechsig (1931) to explain the temperature-dependent exponential decay of thallium-activated potassium chloride. In general phosphors have a distribution of electron traps and the decays due to each depth of trap must be summed over the range of trap depths. The general expression for the decay of phosphorescence due to electron traps is therefore

$$I = \int_0^\infty N_E s \exp(-E/\boldsymbol{k}T)\exp\{-st\exp(-E/\boldsymbol{k}T)\}\,dE \qquad \ldots\ldots(13)$$

where $N_E dE$ is the number of electron traps of depths between E and $E+dE$. If the traps are not saturated then N_E is replaced by n_E representing the number of traps of depth E filled by excitation (Garlick and Gibson 1948b). For special

cases equation (13) may be solved. Randall and Wilkins have given the following solutions:

(*a*) For N_E constant for all trap depths from zero to infinity:

$$I = N_E \boldsymbol{k} T/t. \qquad \ldots\ldots(14)$$

(*b*) For an exponential trap distribution, that is $N_E = a \exp(-\alpha E)$

$$I = \text{constant}/t^{(\alpha \boldsymbol{k} T+1)}. \qquad \ldots\ldots(15)$$

In both cases the decay equation applies to the phosphorescence at times greater than a few microseconds ($st \gg 1$). Many phosphors show this form of decay at room temperature, usually that given by equation (15). The following table gives values of the index $(\alpha \boldsymbol{k} T + 1)$ obtained from measurements of the phosphorescence decay and from the form of the thermoluminescence curve corresponding to the decay.

Values of $(\alpha \boldsymbol{k} T + 1)$ derived from Phosphorescence and Thermoluminescence

Phosphor	ZnS–Cu	ZnS–Cu	ZnS–Cu–Ag	ZnS–CdS–Cu	ZnS–Cu
Phosphorescence ..	1·5	1·38	1·32	1·65	2·0
Thermoluminescence	1·57	1·31	1·30	1·57	2·0

We see that a value of two for the index, previously explained by a bimolecular hypothesis, can be explained by the particular trap distribution of a phosphor specimen.

(*c*) Complex trap distributions: Although equation (13) cannot be solved for such distributions, as shown in Figure 6, these distributions can be correlated with the phosphorescence characteristics by a more approximate method due to Randall and Wilkins. From equation (5) we see that an electron spends a mean time t in traps of depth E given by

$$t = \{\exp(E/\boldsymbol{k} T)\}/s. \qquad \ldots\ldots(16)$$

Thus at time t during decay of phosphorescence for a complex trap distribution the intensity I is mainly due to electrons which spend a mean time t in traps and to the number in such traps, N_t. Thus we may write It proportional to N_t and, from equation (16), E proportional to $\ln t$. Thus by plotting graphically It against $\ln t$ we obtain a curve whose form should correspond to that of the thermoluminescence curve. The success of this approximate method for correlating phosphorescence and thermoluminescence is shown by the curves of Figure 7 due to Wilkins (unpublished studies).

(iii) *Retrapping of Electrons during Phosphorescence and Thermoluminescence*

Randall and Wilkins gave certain experimental results to justify their assumption that retrapping was negligible even in phosphorescence and thermoluminescence processes in photoconducting phosphors such as zinc sulphide. Later workers have provided more conclusive experimental support for their assumption. We shall discuss this below and also the modification of the theory of phosphorescence and thermoluminescence which would be necessary if retrapping occurred. One of the most conclusive experiments showing the absence of retrapping of electrons in zinc sulphide phosphors is that due to Garlick and Wilkins (1948). They used the phosphor giving curve A of Figure 6. The phosphor is first

excited by intense ultra-violet light at a low temperature so that all except the very shallow traps are filled with electrons. The thermoluminescence curve is then measured, as shown in curve A of Figure 8. After this emptying of traps the phosphor is cooled in the dark and then excited by a short weak pulse of ultra-violet light which only fills a few traps with electrons. On warming in the dark curve B is obtained. This curve is not very different in form from curve A except that the area beneath it, representing the total number of trapped electrons, is only 1/1,300 of the area beneath curve A. If retrapping occurred then the few electrons from shallow traps would be retrapped in deeper traps and would not contribute to thermoluminescence at the lower temperatures. The expected curve in this case would be something like curve C. The fact that the measured curve is not very different from that for saturation of the traps shows that no appreciable retrapping could have occurred. Garlick and Gibson (1948b)

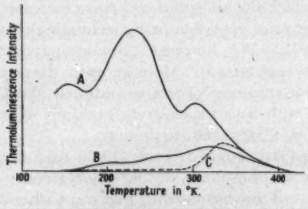


Figure 8. Thermoluminescence curves for a ZnS–Cu phosphor which show the absence of retrapping.
A. Phosphor strongly excited at 90° K. B. Phosphor excited by weak, short pulse of ultra-violet radiation after heating to 500° K. to empty all traps.
C. Theoretical curve expected when retrapping is present.
N.B.—Area under curve A is 1300 × area under curve B.

have made more quantitative experiments to show the absence of retrapping in other sulphide and silicate phosphors. We may include here their basic equations for the luminescence process when the probability of retrapping of an electron in an empty trap is the same as that for capture by an empty luminescence centre:

$$I = -dn/dt = (n^2 s/N_E) \exp(-E/\boldsymbol{k}T). \qquad \ldots\ldots(17)$$

These are for a single depth of trap and should be compared with equations (6). The corresponding expression to equation (10) for the thermoluminescence is

$$I = \{n^2 s \exp(-E/\boldsymbol{k}T)\}/N_E\{1 + (n_E/N)\int_0^T (s/\beta) \exp(-E/\boldsymbol{k}T)\,dT\}^2 \qquad \ldots\ldots(18)$$

and for the phosphorescence

$$I = \{n_E^2 s \exp(-E/\boldsymbol{k}T)\}/N_E\{1 + (n_E/N_E)st \exp(-E/\boldsymbol{k}T)\}^2 \qquad \ldots\ldots(19)$$

with the usual notation. This last equation has a bimolecular form. A more general treatment of phosphorescence due to traps when retrapping occurs has recently been given by Klasens and Wise (1946).

If a phosphor is heated to empty its traps of electrons and then excited, its fluorescence will increase with time in a manner which can indicate the presence of retrapping. If the latter occurs the rise of intensity with time gives a curve having an inflection. The presence of retrapping at high temperature has been investigated in this way by Garlick and Wilkins (1945).

(iv) *Phosphorescence of Short Duration due to Electron Traps*

It has been shown that even at short decay times, of the order of milliseconds, the phosphorescence of copper and silver activated phosphors is due to the action of electron traps. It has also been shown that, except at high temperatures, there is no simple bimolecular mechanism of decay in these phosphors at such times (Garlick and Wilkins 1945). At high temperatures, where the luminescence efficiency is decreasing rapidly with temperature, effects due to the retrapping of electrons have been observed.

(v) *The Electrical Characteristics of Electron Traps*

It has been assumed in the past that the space charge effects observed after the initial rise of the primary photocurrent in phosphors and other insulators may be due to electron traps (Mott and Gurney 1940). Investigations have been made of the way in which the phosphor resistance increases with time as the traps are emptied by heating the phosphor or by irradiating it with infra-red radiation (Reimann 1937, Gisolf 1939); however, no quantitative experimental results of significance have been obtained. More recently Herman and Meyer (1946), using the concepts of trapping processes outlined above, have calculated the change of resistance with time for different numbers of traps initially filled and for different intensities of infra-red irradiation.

In recent studies Garlick and Gibson (1947) have measured the changes in dielectric properties of phosphors during fluorescence, phosphorescence and thermoluminescence. They have shown that traps when filled constitute highly polarizable systems in the phosphor and that in spite of their small concentrations in most matrices (1 in 10,000 lattice atoms) they can give rise to the observed increase in dielectric constant. This is of the same order as the value for the unexcited phosphor. Their experiments show that the magnitude of the changes in dielectric constant and dielectric loss during luminescence depend on the total number of trapped electrons in the phosphor, on the temperature and, to a lesser degree, on the distribution of trap depths in a given material. The large polarizability would be expected from the relatively small binding energies of the trapped electrons (0·1 to 1·0 ev.) (Mott and Gurney 1940). The filled traps have a relaxation time of about 10^{-7} sec. which gives rise to a maximum in the loss factor of the excited phosphor at radio frequencies of a few megacycles per second. No such effect is observed for the unexcited phosphor. Garlick and Gibson have shown that the change in dielectric loss of phosphors, previously attributed to the presence of electrons in the conduction band (Gisolf 1939, de Groot 1946), is due almost entirely to trapped electrons. There is a probability that the primary photocurrent in phosphors, which decays so rapidly after the initial excitation, is in fact a displacement current associated with the creation of the highly polarizable centres formed by trapped electrons. It may be significant in this respect that the excitation spectra for luminescence, photoconductivity and the dielectric changes are very similar for a given phosphor specimen.

(vi) *The Nature of the Electron Traps in Phosphors*

In some solids the nature of the electron traps is well known. For example, the structure of F, U and V centres in the alkali halides has been determined (Seitz 1946). In thallium activated alkali halide phosphors the traps are actually

metastable states due to pairs of thallium ions (Seitz 1938). However, in most photoconducting phosphors little is known about the nature of the electron traps. We may summarize what is known so far that might help towards an interpretation of trapping states in such phosphors: (*a*) Retrapping of electrons is not usually observed. This means that electrons from traps return to empty centres which are relatively close to the traps. Thus traps may be formed in the immediate region of the centres (Garlick and Gibson 1948b). (*b*) Some activator impurities seem to give rise to particular trap distributions which would appear to support the idea in (*a*) above (Garlick 1948). (*c*) In zinc sulphide phosphors the longest afterglow is produced by copper as activator. The same trap distribution is obtained by introducing copper into a lattice at temperatures as low as 300° C. as is obtained for phosphors made entirely at high temperatures with copper impurity, provided that in the first case the matrix lattice was first formed at high temperatures (Garlick 1948). (*d*) The electron trap distribution alters in a regular manner as the matrix lattice of a zinc sulphide phosphor is changed by adding increasing amounts of cadmium sulphide in preparation (Garlick 1948). (*e*) The total number of traps is usually of the same order as the number of activator impurity centres in a phosphor (Randall and Wilkins 1945).

These facts are little more than pointers to the nature of traps in phosphors. More careful studies are needed with stricter control of phosphor preparation and removal of all traces of unwanted impurities. We shall refer later to some specific effects of impurities on trap formation in silicate phosphors (§ 4 (iii)).

§ 4. LUMINESCENCE AND PHOSPHOR CONSTITUTION AND PREPARATION

In this section we shall consider two main classes of phosphors, the sulphides and the oxygen-containing phosphors, such as silicates and tungstates. We first discuss zinc and zinc–cadmium sulphide phosphors.

(i) *Zinc and Zinc–Cadmium Sulphide Phosphors*

The general methods for obtaining the basic materials for phosphor preparation have been described by various workers (Ward 1948, Kröger 1948, Leverenz 1945). The importance of a high standard of purity has been stressed by Leverenz (1945, 1948). The most significant investigations showing the effect of phosphor preparation and constitution on the luminescence characteristics have been those of Rothschild (1946), Wells (1944), Kröger and Hellingmann (1948), and Leverenz (1945). The basic constituents of zinc sulphide phosphors before firing are usually: (*a*) the matrix material: ZnS, CdS or ZnS–CdS; (*b*) the activating impurity: Cu, Mn, Ag, or excess Zn, Cd; (*c*) a flux material: a few percent of an alkali halide (not a fluoride). We shall not discuss the manganese-activated phosphors here. The mixed systems formed by zinc, cadmium and manganese sulphides have been thoroughly studied by Kröger (1940a) and it is almost certain that the activator occupies a substitutional site in the matrix. Our attention is confined to copper, silver or self-activated phosphors. Previously it was thought that the luminescence of pure zinc sulphide was due to centres formed by excess zinc in interstitial lattice sites. Analogies were assumed to exist with the characteristics of pure zinc oxide phosphors with excess zinc present. However, in recent investigations Rothschild (1946) and Kröger and Hellingmann (1948) have shown that the blue luminescence of pure zinc sulphide

does not occur if the phosphor is made in the absence of halide ions (exclusive of fluorine). Such inclusions are not necessary to form green emitting zinc oxide. It has long been recognized that a few per cent of an alkali halide added to a zinc sulphide phosphor before firing improves the crystallinity and efficiency of the final product. However, Wells (1944) has shown that the amount necessary for this is only of the same order as the activator concentration. It now appears from the work of Rothschild that the halide is an active constituent in the structure of the emitting centres.

Kröger and Hellingmann (1948) have found that in the absence of halide ions the pure zinc sulphide phosphors only exhibit the low temperature edge emission shown in Figure 5. When these phosphors are refired with inclusion of a halide or in an atmosphere of hydrogen sulphide or hydrogen and hydrogen chloride the characteristic blue emission develops at the expense of the edge emission. This fall in the edge emission is attributed to the capture of positive holes by blue emitting centres so that electrons excited into the conduction levels can only recombine with holes in the blue emitting centres. The development of the blue emission is accompanied by an increase of absorption in the "tail" of the fundamental absorption band. The above workers conclude from their extensive experimental studies that the emission centres in copper, silver or self-activated phosphors are formed by halides of the impurity activator existing as molecular groups in the matrix lattice. It is still not known whether the impurity metal occupies substitutional or interstitial lattice sites. A concentration of copper in zinc sulphide which exceeds 10^{-4} per ZnS mole causes the characteristic green emission for copper activation to diminish and a new band in the blue region to appear. Rothschild considers this new emission to be due to copper and not to a reversion to the self-activated emission. The fact that copper can activate a zinc sulphide lattice at firing temperatures as low as 300° C. (see § 3 above) is not considered to be a conclusive proof that it occupies interstitial sites (Kröger and Hellingmann 1948).

(ii) *The Alkaline Earth Sulphide Phosphors*

The knowledge of alkaline earth sulphide phosphors, activated by bismuth and other metals, previous to 1939 has been reviewed by Birus (1942) and by Rupp (1937). We shall discuss some recent investigations of the preparation conditions of these phosphors due to Wells (1944) and their thermoluminescence characteristics (Randall and Wilkins 1945, Wells 1944). The usual method of making these phosphors is by heating the alkaline earth carbonate or oxide with sulphur, a suitable amount of the activator and a flux. Wells has found that after firing the phosphor contains sulphide and sulphate and probably some complexes of the cation. If the flux is omitted then very little sulphide is formed and the product is only feebly luminescent. If prepared by heating the carbonate in a stream of nitrogen and carbon bisulphide almost complete conversion to sulphide occurs but in the absence of a suitable flux no luminescence is observed in the final product. For calcium and strontium sulphide phosphors the most effective fluxes seem to be the oxysalts of sodium and it is found that the cation Na^+ is the important factor. In the case of barium sulphide potassium salts are more effective. When the activator impurity is manganese or a rare earth the flux seems to be less important. This is presumably because such activators already possess in themselves conditions favourable to luminescence emission.

Bismuth-activated calcium and strontium sulphides exhibit a very long afterglow at room temperature due to the large numbers of deep electron traps which they contain (0·7–1·0 ev.). Figure 9 gives thermoluminescence curves for a calcium sulphide phosphor activated by bismuth. Curve A is that for the phosphor fired without flux while curve B is that for the refired phosphor, sodium peroxide having been included as a flux. The trap distributions in such phosphors

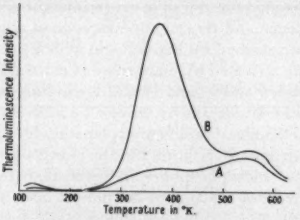


Figure 9. Thermoluminescence curves of CaS–Bi phosphors.
A. CaS–Bi made at 900°c. without flux.
B. Phosphor A reheated at 900 with Na_2O_2 (10% by weight).

are characteristic of the activating impurity. The alkaline earth sulphide phosphors have phosphorescence and thermoluminescence properties similar to those of zinc sulphide phosphors. As in the latter phosphors, there is no evidence of appreciable retrapping of electrons during these processes. The phosphorescence efficiency of these materials can in some cases be improved by adding a second activator, notably samarium (Rothschild 1936). Secondary activators play an important part in the sensitization of phosphors of this class to the stimulating effects of infra-red radiation which we shall discuss in the next section.

(iii) *Silicates and other Oxygen-containing Phosphors*

(*a*) *Silicate phosphors.*

Zinc and zinc beryllium silicate phosphors activated by manganese have been studied by many workers because of their practical importance. Investigations of more fundamental interest include those of Kröger (1940a), Fonda (1939, 1941), Froelich (1943), Leverenz (1945, 1948), and Schulman (1946). Froelich (1948), has recently described methods of obtaining the basic materials for these phosphors in a high state of purity. In these phosphors it is thought that the manganese activator occupies substitutional positions in the lattice but some evidence for interstitial manganese is reviewed below.

The only real compound of zinc oxide and silica has been found to be the orthosilicate Zn_2SiO_4 (Bunting 1930) which is isomorphous with beryllium silicate and manganese orthosilicate. The ternary system formed by these three compounds has been extensively studied by Kröger (1940a). When excess silica is present the emission spectrum changes, as we have seen above, but no new compound is formed. When the silica excess is high x-ray analysis reveals a crystobalite structure. Schulman (1946) has recently attempted to relate the emission spectrum of silicate phosphors activated by manganese to their structure and to the coordination of the Mn^{2+} activator ions. He has followed the ideas of Linwood and Weyl (1942) who have associated the different spectral emissions

of manganese centres in glasses with the different coordination numbers of the manganese ions. The red emission of zinc silicate phosphors containing high concentrations of activator at low temperatures and of phosphors containing excess silica is attributed to manganese centres having a coordination number higher than four (the value in green emitting zinc silicate). Schulman suggests that in these cases, and also in the case of red emission from zinc beryllium silicates, the manganese ions responsible are in interstitial positions.

The phosphorescence and thermoluminescence of silicate phosphors due to electron traps has been studied by Randall and Wilkins (1945). The phosphorescence of zinc silicate activated by manganese is enhanced if arsenic is included in the phosphor (Froelich and Fonda 1942). Schulman suggests that As^{5+} ions replace Si^{4+} ions in the lattice and so leave an excess charge which behaves as a trap for electrons. He suggests a similar explanation for the same effect produced by inclusion of tin in zinc beryllium silicate (Leverenz 1946). The thermoluminescence curves of such phosphors are given in Figure 10. The variation of their form with arsenic or tin content has not yet been studied.

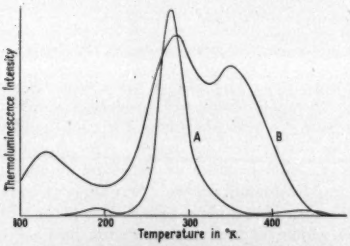


Figure 10. Thermoluminescence curves for long afterglow silicate phosphors. A. Zn_2SiO_4–Mn–As phosphor. B. $ZnBeSiO_4$–Mn–Sn phosphor made by Leverenz.

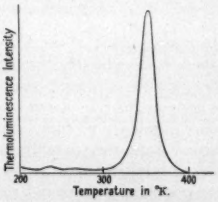


Figure 11. Thermoluminescence curve for a $SrSiO_3$–Eu phosphor excited by 2537 A. (Garlick and Gibson).

Other silicate phosphors of interest include those of the alkaline earths activated by europium (McKeag and Ranby 1942). In particular strontium silicate activated by europium shows unique luminescence characteristics (Garlick and Gibson 1948b). It exhibits a broad spectral emission band in the blue-green region and is excited by radiation of 3650 A. or that of shorter wavelength. The luminescence decay after excitation by 3650 A. radiation is exponential with time, its decay rate being independent of temperature and characteristic of a forbidden optical transition in the europium centres. The half life of the decay is about 2·3 seconds which is of a unique magnitude for all but organic molecules. Excitation by 2537 A. radiation also produces this decay but in addition gives rise to an exponential decay of much longer duration whose rate varies exponentially with temperature. It is due to a trapping state of depth 0·75 ev. The thermoluminescence curve of this phosphor is given in Figure 11. It indicates the single depth of trap deduced from the phosphorescence characteristics. Photoconductivity is exhibited by this type of phosphor when excited by 2537 A. radiation but, as in zinc and zinc beryllium silicate phosphors, it appears to be characteristic of the pure lattice and shows little correlation with the luminescence (Hill and Aronin 1940).

Barium silicate activated by lead has been investigated by Butler (1947). The luminescence is blue or blue-green depending on the lead content and on the lattice structure. The metasilicate shows the most efficient luminescence. The latter occurs even for very high lead concentrations. Butler has also found that an excess of silica is essential to the emission efficiency. Calcium silicate activated by lead and manganese has been studied by Merrill and Schulman (1948). It can be stimulated to give emission when irradiated with infra-red radiation or long wavelength light. The effect has a maximum for 1 μ radiation.

(*b*) *Tungstate phosphors.*

The emission centres in pure tungstate phosphors are thought to be WO_4 ions. Recent studies of pure tungstates have been made by Hill (1946), Fonda (1944) and Kröger (1947a). We have already referred to the effect of low temperatures on the luminescence efficiency of these phosphors. In other studies Kröger *et al.* (1947) have shown that the luminescence efficiency of tungstates improves with increase in crystal perfection. Tungstates may also be activated by impurities, such as samarium and uranium. The latter activator produces phosphorescence of long duration in the phosphor due to traps (Garlick 1948).

(*c*) *Zinc oxide and aluminium oxide phosphors.*

The emission spectra of zinc oxide phosphors have been measured by Randall (1939) at low temperatures and more recently by Leverenz (1945), and by Nicoll (1948) who has studied their variation with temperature. Zinc oxide exhibits an ultra-violet "edge emission" of form similar to that of zinc sulphide at low temperatures. At room temperature and under cathode-ray excitation a narrow spectral band appears in the emission with a maximum at about 3800 A. When the phosphor is prepared in reducing atmospheres of carbon monoxide a strong green emission is produced. Other reducing atmospheres do not seem to be as suitable and it is probable that carbon plays a specific part in the luminescence centres of the phosphors of high efficiency.

The luminescence of ruby (Al_2O_3–Cr), first studied by Deutschbein (1932), has recently been investigated by Thosar (1938) and its decay measured over a wide range of temperatures by Garlick and Wilkins (1945). The decay is characteristic of forbidden transitions in the Cr^{3+} ions which have been shown by Thosar to replace Al^{3+} in the matrix lattice.

(iv) *Tetravalent Manganese as a Luminescence Activator*

In recent studies Kröger (1947b) has shown that tetravalent manganese in magnesium titanate exhibits a red emission whose spectrum has a detailed structure similar to that of emission from Cr^{3+} ions in the same matrix. Since the two activator ions manganese and chromium have the same electronic structure in the tetravalent and trivalent states respectively such similarity would be expected.

§5. INFRA-RED SENSITIVE PHOSPHORS

By infra-red sensitive phosphors we mean those whose afterglow is enhanced by irradiation with long wavelength light or infra-red radiation. This effect is known as stimulation and is not very marked in conventional sulphide phosphors. During the late war zinc sulphide and alkaline earth sulphide phosphors were developed containing two activators and having relatively large infra-red stimulability. Their stimulation was used as a means of detecting infra-red radiation

of wavelengths around 1μ. The increased sensitivity to infra-red radiation is due to the secondary activator while the emission is usually characteristic of the first or dominant activator, the latter having a higher concentration than the secondary activator. The characteristics of these phosphors which have been investigated because of their fundamental interest may be summarized as follows:

(*a*) The stimulated emission spectrum.

(*b*) The stimulation spectrum, that is, the spectral response of the stimulation in the infra-red.

(*c*) The decay of the stimulated emission with time and its dependence on the infra-red intensity.

(*d*) The relations between thermoluminescence (i.e. thermal ejection of trapped electrons) and stimulation (i.e. the optical ejection of trapped electrons).

The important secondary activator in zinc sulphide phosphors is found to be lead. These phosphors were developed by Fonda (1946) and studied also by Urbach *et al.* (1948). The alkaline earth sulphides with samarium as the secondary activator of most interest have been investigated by Urbach *et al.* (1946) and by Ward (1946). We give below a brief résumé of their characteristics, these phosphors having been surveyed elsewhere in more detail (Garlick 1947). At room temperature the emission spectra of stimulated zinc sulphides, containing copper and lead, manganese and lead or copper and manganese as activator pairs, consist of bands due to both impurities but one activator emission predominates. The stimulation spectra of zinc sulphides containing lead as secondary activator consist of two broad bands with maxima at $0{\cdot}8\mu$ and $1{\cdot}3\mu$ respectively at ordinary temperatures. Fonda has reported an exponential decay of the stimulated emission with time but other workers have found hyperbolic decays for such phosphors (Urbach *et al.* 1948). There have been no extensive studies of the variation in decay characteristics with temperature or infra-red intensity and the basic processes still remain obscure.

The principal alkaline earth sulphides having infra-red stimulability are those containing cerium, manganese or europium as dominant activators and samarium as secondary activator. The emission spectra at room temperature are characteristic of the dominant activator only while the stimulation spectrum is characteristic of samarium and has a single band with maximum at about 1μ. The stimulated emission decays hyperbolically with time and the following empirical formula has been suggested for it by Ellickson and Parker (1946, 1948):

$$I_t = I_0/(1+\alpha t)^n, \qquad \ldots\ldots (20)$$

where n is a constant and is about 1·25 for many phosphors and α is another constant which depends on the infra-red intensity. Ellickson and Parker have suggested that the hyperbolic decay is due to a summation of elementary bimolecular processes in their thick phosphor layers for each successive layer beneath the irradiated surface. They introduce absorption of infra-red and emitted light to explain the value of the index n. Their assumptions have not been supported by absorption measurements and have been shown to be incorrect by the occurrence of the same law (equation (20)) for thin phosphor layers (Urbach *et al.* 1948). They have shown by experiment that the stimulation spectrum decays as a whole with the decay of emission, even when monochromatic infra-red radiation is used.

Theoretical Studies of Infra-red Stimulation

Attempts to provide a theoretical explanation of infra-red stimulation and quenching of emission have been made by various workers. According to Brauer (1947) the secondary activator provides the electron traps but its normal emission is suppressed because infra-red radiation not only ejects the trapped electrons but also raises electrons from the filled band of the matrix into the ground states of the secondary activator centres. This prevents recombination of excited electrons with these centres. This explanation is attractive and is similar to that described above for thermal quenching of luminescence. It lacks any definite experimental support and hardly applies to zinc sulphide phosphors at low temperatures where the stimulation spectrum is relatively independent of the particular activators present (author's unpublished data). Urbach (1948) has recently discussed the above infra-red effects and has stressed the difficulties in establishing a theory for them. In particular he points out that the previous assumption, that the optical and thermal activation energies for electron traps should be related by the Franck–Condon principle, does not hold. The stimulation spectra of phosphors appear to bear little relation to their thermoluminescence characteristics, the latter representing the trap distributions.

So far the experimental data for infra-red sensitive phosphors and for more conventional materials is too inadequate to establish a precise theoretical model. Simultaneous studies of thermal and optical ejection of trapped electrons over wide temperature ranges are required in order to elucidate the basic mechanisms of stimulation and quenching by long wavelength radiation.

§ 6. CONCLUSION

In the above survey it has not been possible to cover the whole field of luminescence. We therefore give here an indication of reviews of other aspects not dealt with above. The characteristics of phosphors under electron bombardment will be reviewed in the near future (Garlick, in press). Bowen (1947), has recently provided a discussion of luminescence and related phenomena in organic molecules and crystals. Destriau (1947) has described the characteristics of the luminescence produced in zinc sulphide phosphors by applied electric fields.

Very recently phosphors have been used in conjunction with electron photomultipliers as scintillation counters for elementary particles and gamma rays. It is to be hoped that some effort will be devoted to a study of the fundamental aspects of particle excitation with the high sensitivity multipliers now available.

ACKNOWLEDGMENTS

Acknowledgments are due to those authors and publishers who have given permission for the reproduction of diagrams from their papers; to Drs. M. H. F. Wilkins, A. F. Wells and A. F. Gibson for their permissions to quote from unpublished studies made in the author's laboratory.

REFERENCES

Birus, K., 1942, *Ergebn. exact. Naturw.*, **20**, 183.
Bloch, F., 1928, *Z. Phys.*, **52**, 555.
Bowen, E. J., 1947, *Quart. Rev. Chem. Soc.*, **1**, 1.
Brasefield, C. J., 1940, *Phys. Rev.*, **57**, 162; **58**, 436.

BRAUER, P., 1947, *Z. Naturforsch.*, **2a**, 238.
BÜNGER, W., and FLECHSIG, W., 1931, *Z. Phys.*, **67**, 42.
BUNTING, E. N., 1930, *Bur. Stand. J. Res., Wash.*, **4**, 131.
BUTLER, K. H., 1947, *Trans. Electrochem. Soc.*, **91**, 27.
DESTRIAU, G., 1947, *Phil. Mag.*, **38**, 700.
DEUTSCHBEIN, O., 1932, *Ann. Phys., Lpz.*, **14**, 172.
ELLICKSON, R. T., and PARKER, W. L., 1946, *Phys. Rev.*, **70**, 290; 1948, *Solid Luminescent Materials*, Paper 19, p. 327 (New York : John Wiley).
FARADAY SOCIETY, 1939, *Symposium on Luminescence*.
FONDA, G. R., 1939, *J. Phys. Chem.*, **43**, 561; 1940, *Ibid.*, **44**, 851; 1941, *Ibid.*, **45**, 282; 1944, *Ibid.*, **48**, 303; 1946, *J. Opt. Soc. Amer.*, **36**, 382.
FRENKEL, J., 1931, *Phys. Rev.*, **37**, 17.
FROELICH, H. C., 1943, *J. Phys. Chem.*, **47**, 669; 1948, *Solid Luminescent Materials*, Paper 3, p. 44 (New York : John Wiley).
FROELICH, H. C., and FONDA, G. R., 1942, *J. Phys. Chem.*, **46**, 1.
GARLICK, G. F. J., 1947, *Electronics Forum*, **7**, 5; 1948, *Solid Luminescent Materials*, Paper 5, p. 87 (New York : John Wiley); (In press) *Advances in Electronics*, **2** (New York : Academic Press).
GARLICK, G. F. J., and GIBSON, A. F., 1947, *Proc. Roy. Soc.* A, **188**, 485; 1948 a, *Nature, Lond.*, **161**, 359; 1948 b, *Proc. Phys. Soc.*, **50**, 574.
GARLICK, G. F. J., and WILKINS, M. H. F., 1945, *Proc. Roy. Soc.* A, **184**, 408; 1948, *Nature, Lond.*, **161**, 565.
GISOLF, J. H., 1939, *Physica*, **6**, 918.
GISOLF, J. H., and KRÖGER, F. A., 1939, *Physica*, **6**, 1101.
DE GROOT, W., 1946, *Physica*, **12**, 402.
DE GROOT, W., GISOLF, J. H., and KRÖGER, F. A., 1941, *Physica*, **8**, 805.
HERMAN, R. C., and MEYER, H. F., 1946, *J. Appl. Phys.*, **17**, 743.
HILL, A. G., and ARONIN, L. R., 1940, *Phys. Rev.*, **57**, 1090.
HILL, C. G. A., 1946, *Trans. Faraday Soc.*, **42**, 685.
VON HIPPEL, A., 1936, *Z. Phys.*, **101**, 680.
JOHNSON, R. P., 1939, *J. Opt. Soc. Amer.*, **29**, 387.
KLASENS, H. A., 1946, *Nature, Lond.*, **158**, 306.
KLASENS, H. A., and WISE, M. E., 1946, *Nature, Lond.*, **158**, 483; 1948, *J. Opt. Soc. Amer.*, **38**, 226.
KRÖGER, F. A., 1939, *Physica*, **6**, 674; 1940 a, *Luminescence in Solids containing Manganese* (Amsterdam : Van Campen); 1940 b, *Physica*, **7**, 92; 1947 a, *Nature, Lond.*, **159**, 674; 1947 b, *Ibid.*, **159**, 706; 1947 c, *Some Aspects of the Luminescence of Solids* (Amsterdam : Elsevier).
KRÖGER, F. A., and HELLINGMANN, J. E., 1948, *J. Electrochem. Soc.*, **93**, 156.
KRÖGER, F. A., HOOGENSTRAATEN, W., BOTTEMA, M., and BOTDEN, TH. P. J., 1948, *Physica*, **14**, 181.
KRÖGER, F. A., *et al.*, 1947, *Philips Res. Rep.*, **2**, 177, 183.
LEVERENZ, H. W., 1945, *Final Report on Radar Indicator Screens*, PB-25481 (Washington); 1946, *R.C.A. Rev.*, **7**, 199; 1948, *Solid Luminescent Materials*, Paper 8, p. 148 (New York : John Wiley).
LINWOOD, S. H., and WEYL, W. A., 1942, *J. Opt. Soc. Amer.*, **32**, 443.
MCKEAG, A. H., and RANBY, P. W., 1942, Brit. Pat. 544, 160.
MERRILL, J. B., and SCHULMAN, H. J., 1948, *J. Opt. Soc. Amer.*, **38**, 471.
MOTT, N. F., and GURNEY, R. W., 1939, *Trans. Faraday Soc.*, **35**, 69 ; 1940, *Electronic Processes in Ionic Crystals* (Oxford : University Press).
NICOLL, F. H., 1948, *J. Opt. Soc. Amer.*, **38**, 817.
PEIERLS, R., 1932, *Ann. Phys., Lpz.*, **13**, 905.
RANDALL, J. T., 1939, *Trans. Faraday Soc.*, **35**, 2.
RANDALL, J. T., and WILKINS, M. H. F., 1939, *Rep. Prog. Phys.*, **6**, 174 (London : Physical Society) ; 1945, *Proc. Roy. Soc.* A, **184**, 366.
REIMANN, A. L., 1937, *Nature, Lond.*, **140**, 501.
RIEHL, N., 1939 a, *Trans. Faraday Soc.*, **35**, 135; 1939 b, *Z. Phys.*, **114**, 682; 1939 c, *Z. tech. Phys.*, **20**, 152.

ROTHSCHILD, S., 1936, *Z. Phys.*, **87,** 757; 1937, *Ibid.*, **108,** 24; 1946, *Trans. Faraday Soc.*, **42,** 638.

RUPP, H., 1937, *Die Leuchtmassen und ihre Verwendung* (Berlin : Borntraeger).

SCHÖN, M., 1942, *Z. Phys.*, **119,** 470.

SCHULMAN, J. H., 1946, *J. Appl. Phys.*, **17,** 902.

SEITZ, F., 1938, *J. Chem. Phys.*, **6,** 150; 1939, *Trans. Faraday Soc.*, **35,** 74; 1946, *Rev. Mod. Phys.*, **18,** 384.

THOSAR, B. V., 1938, *Phil. Mag.*, **26,** 380.

URBACH, F., 1948, *Solid Luminescent Materials*, Paper 6, p. 115 (New York : John Wiley).

URBACH, F., PEARLMAN, D., and HEMMENDINGER, H., 1946, *J. Opt. Soc. Amer.*, **36,** 373; 1948, *Solid Luminescent Materials*, Paper 16, p. 279 (New York : John Wiley).

URBACH, F., PEARLMAN, D., and NAIL, N. R., 1948, *Solid Luminescent Materials*, Paper 21, p. 358 (New York : John Wiley).

URBACH, F., URBACH, A., and SCHWARZ, M., 1947, *J. Opt. Soc. Amer.*, **37,** 122.

WARD, R., 1946, *J. Opt. Soc. Amer.*, **36,** 351; 1948, *Solid Luminescent Materials*, Paper 2, p. 22 (New York : John Wiley).

WELLS, A. F., 1944, *Ministry of Home Security Report*, RC(C) 119 (available at D.S.I.R. Library).

N.B.—Much of the recent work of Kröger and his associates at Eindhoven has been described in *Some Aspects of the Luminescence of Solids*, by F. A. KRÖGER, 1947 (Amsterdam : Elsevier).

SPONTANEOUS FLUCTUATIONS

By D. K. C. MacDONALD
Clarendon Laboratory, Oxford

CONTENTS

PAGE
§1. Introduction 56
§2. Fluctuations analysis 56
§3. Survey of development in recent years 69
3.1. Electrical noise 69
3.2. Fundamental valve noise problems 73
3.3. Low-frequency noise in valves and semiconductors, etc. 75
3.4. Spontaneous fluctuations of temperature and radiation 78

§1. INTRODUCTION

THE writer has approached the task of composing this review with no little trepidation. Apart from a brief summary of then current development (Brown 1938) of the Brownian movement in mechanical systems, no precedent exists in this series of reviews to guide him in this subject. The influence of fluctuation phenomena—frequently called "noise" in certain fields—seems to extend so far that it is difficult to know where the subject ends and physics begins!

In view of the general title of these publications it seems somewhat inappropriate to regress as far as 1827 when the biologist Robert Brown first observed the irregular continual motion of minute pollen grains in water which has given the name *Brownian movement* to the whole genus of related random phenomena. Furthermore, a number of general reviews exist today (e.g. de Haas–Lorentz 1913, Fürth 1920a, b, Barnes and Silverman 1934, MacDonald 1948b, see also Fürth's edition of Einstein's papers (1926)) which contain details of the earlier history of the Brownian movement. The present paper has therefore been divided into two main parts: first, a selected survey of some developments of fluctuations analysis most frequently met with today, but omitting any reference to the more general problems of statistical analysis to be found in standard works; second, a summary of general development in fluctuation phenomena covering more or less the past decade, omitting again such fundamental studies as the application of statistical mechanics to physical chemistry and thermodynamics.

The writer, in advance, asks indulgence for the omission of topics of interest which he feels, in such a subject, must be numerous.

§2. FLUCTUATIONS ANALYSIS

The phenomena of osmosis and diffusion were of course well known to Einstein (1905 etc.) and he recognized in effect that they were simply consequences of the random thermal motion of molecules when unequal densities are present in different parts of the system. If a macroscopically uniform concentration is finally attained, then this is interpreted as only a statistical equilibrium inasmuch as, on the average, as many molecules will be transported through any given surface in one direction as in the opposite one.

The density distribution $f(r, t)$ predicted by the solution of the diffusion equation is then regarded as a probability function for the behaviour of any

individual molecule. The mean square fluctuation of the distance travelled, for example, by a molecule may then be deduced. In the classical case of the Brownian movement of minute particles, the associated diffusion equation in one dimension is

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2}, \qquad \ldots\ldots(1)$$

where $f(x, t)$ gives the relative density of particles at x at time t, and D, by definition, is the diffusion constant. By consideration of a specific case the value of D for a given system may be deduced. For spherical particles, radius a, suspended in a fluid of viscosity η, one finds

$$D = \frac{\boldsymbol{k}T}{6\pi\eta a}, \qquad \ldots\ldots(2)$$

where T is the absolute temperature. If then we consider an initial concentration of particles at $x=0$, $t=0$, the appropriate solution of (1) is of the form

$$f(x,t) = \frac{1}{\sqrt{(4\pi D)}}\frac{\exp(-x^2/4Dt)}{\sqrt{t}}. \qquad \ldots\ldots(3)$$

Interpreting this as a probability function, it follows that the mean square distance travelled by a particle in time t is

$$\overline{x^2} = 2\int_0^\infty \frac{1}{\sqrt{(4\pi Dt)}}\exp(-x^2/4Dt)\,x^2\,dx$$

$$= 2Dt = 2\left(\frac{1}{6\pi\eta a}\right)\boldsymbol{k}Tt. \qquad \ldots\ldots(4)$$

Einstein generalized the derivation to the mean square fluctuation of any observable parameter, θ, of a system in thermal equilibrium

$$\overline{\delta\theta^2} = 2B\boldsymbol{k}Tt, \qquad \ldots\ldots(5)$$

where B is the "mobility of the system with respect to θ"; thus, in the case of the spherical particle, B is the limiting velocity reached under the action of unit applied force, which by Stokes' Law is $1/6\pi\eta a$, thus establishing agreement between (5) and (4). Einstein immediately applied this formula to the fluctuation of electric charge in a conductor of resistance R. In that case B is the current which flows under application of unit electromotive force which, by Ohm's Law, is $1/R$. Thus we have

$$\overline{\delta q^2} = 2\boldsymbol{k}Tt/R, \qquad \ldots\ldots(6)$$

where q is the electrical charge spontaneously transported across any section of the conductor in time t. This expression could not be examined experimentally until some 20 years later (Johnson 1925) when the technique of radio amplification had become sufficiently developed; the phenomenon is frequently called "Johnson noise".

It is not difficult to see a limitation in Einstein's formula since physical intuition suggests that, for sufficiently small t, $\overline{\delta x^2} \sim t^2$.

Another deduction of Einstein's was that the average kinetic energy of a Brownian particle is $\frac{1}{2}m\overline{u^2} = \frac{1}{2}\boldsymbol{k}T$, and again we see that this expression cannot apply completely generally in the case of a particle starting, for example, initially from rest.

Einstein's general analysis suggests closely the later interpretations by Born (1926) and Ehrenfest (1927) of the wave-function derived from Schrödinger's equation. When applied, for example, to the case of a single electron the wave-function is regarded as providing a probability function for the motion rather than a strictly determinate statement of the electron's progress. Fürth (1948a) has drawn attention to the close resemblance between many aspects of quantum analysis and of random fluctuations. In particular we may define an effective velocity of transport of a particle, V say, by the equation

$$V^2 = \overline{x^2}/t^2 = 2B\boldsymbol{k}T/t \qquad \ldots\ldots(7)$$

from (4), and we have then: $\sqrt{x^2}\sqrt{V^2} = 2B\boldsymbol{k}T, \qquad \ldots\ldots(8)$

which may be regarded as the analogue of Heisenberg's equation: $\Delta q \Delta p \simeq \boldsymbol{h}$. Both expressions can be considered as "equations of limiting precision" for systems whose ultimate accuracy of observation is determined either by thermal or quantum effects.

Almost concomitantly with Einstein, van Smoluchowski (1906) published a detailed analysis of the Brownian movement in which essentially the composition of the overall motion from individual collisions was examined (cf. also Fürth 1920b). The treatment basically is analogous to the derivation of the error law by combination of small deviations and leads effectively to the same limiting expression as that of Einstein for "long" periods of observation ($\overline{\Delta x^2} \sim t$). For "short" t, however, another formula compatible now with physical intuition emerges. One may say that this analysis formed the starting-point for the concept of correlation-interval or "probability after-effect" (Wahrscheinlichkeitsnachwirkung) which now plays a leading rôle in much fluctuations analysis.

Similar results were obtained by Langevin (1908) shortly after by an elegant general analytic method. If we consider the case of a free particle of mass m in a viscous fluid then the differential equation of motion is

$$m\frac{du}{dt} + \lambda u = 0, \qquad \ldots\ldots(9)$$

where $\lambda = 6\pi\eta a$ for a spherical particle obeying Stokes' law as before.

To include the influence of molecular bombardment, Langevin proposed that we now write

$$m\frac{du}{dt} + \lambda u = F(t), \qquad \ldots\ldots(10)$$

$F(t)$ being a *rapidly* fluctuating force term whose average is zero but having a definite mean square value. These requirements we may express by setting $\overline{F(t)} = 0$; $\overline{F(t)F(t+\tau)} \simeq 0$, except for $\tau \to 0$; $\overline{F(t)^2} = \Phi$, say. We proceed then to "solve" this equation formally; thus

$$u = u_0 e^{-\beta t} + e^{-\beta t}\int_0^t e^{\beta\xi} A(\xi)\, d\xi, \qquad \ldots\ldots(11)$$

$$\beta = \lambda/m; \qquad A(t) = F(t)/m.$$

More properly we should say that the statistical properties of $u - u_0 e^{-\beta t}$ are the same as those of $e^{-\beta t}\int_0^t e^{\beta t} A(\xi)\, d\xi$. Chandrasekhar (1943) has given a detailed study of the analysis of such stochastic equations; see also Niessen and Bakker (1938).

From (11) we have

$$\bar{u} = u_0 e^{-\beta t}, \text{ since evidently } \overline{\int_0^t e^{\beta\xi} A(\xi)\, d\xi} = 0.$$

This immediately indicates that over a period $\sim 1/\beta$ the initial velocity of the particle will influence the observations. We also find

$$\overline{u^2} = u_0^2 e^{-2\beta t} + \frac{\Theta}{2\beta}\{1 - e^{-2\beta t}\}, \qquad \ldots\ldots(12)$$

where $\Theta = \int_{-\infty}^{\infty} \overline{A(\eta)A(\eta+\tau)}\, d\tau$.

For $t \to \infty$, $\overline{u^2} = \Theta/2\beta$; we know, however, that in this case $\frac{1}{2}m\overline{u^2} = \frac{1}{2}\boldsymbol{k}T$ from Einstein's limiting solution. That is,

$$\Theta = \frac{2\beta \boldsymbol{k} T}{m}. \qquad \ldots\ldots(13)$$

Thus for a particle initially at rest ($u_0 = 0$):

$$\overline{u^2} = \frac{\boldsymbol{k}T}{m}\{1 - e^{-2\beta t}\}. \qquad \ldots\ldots(14)$$

Hence for small intervals t, $\overline{u^2} = \boldsymbol{k}T/m$ has to be replaced by

$$\tfrac{1}{2}m\overline{u^2} = \tfrac{1}{2}\boldsymbol{k}T\{2\beta t\} = \beta \boldsymbol{k} T t. \qquad \ldots\ldots(15)$$

We may likewise analyse the behaviour of the coordinate x starting from $dx/dt = u$, and hence: $\beta^2\overline{x^2} = \overline{(u-u_0)^2} + \overline{\left\{\int_0^t A(\xi)\, d\xi\right\}^2} - 2\overline{\left\{u\int_0^t A(\xi)\, d\xi\right\}}$, leading to

$$\beta^2\overline{x^2} = u_0^2\{1 - 2e^{-\beta t} + e^{-2\beta t}\} + \frac{\boldsymbol{k}T}{m}\{-3 + 4e^{-\beta t} - e^{-2\beta t}\} + \frac{2\boldsymbol{k}T}{m}\beta t, \qquad \ldots\ldots(16)$$

using the now known value for Θ. Thus as $t \to \infty$: $\beta^2\overline{x^2} \to 2\boldsymbol{k}T\beta t/m$, i.e. $\overline{x^2} = 2\boldsymbol{k}Tt/m\beta$, $= 2\boldsymbol{k}Tt/\lambda$ in agreement with (Einstein's) equation (4). And for $t \to 0$: $\beta^2\overline{x^2} = u_0^2\beta^2 t^2$, i.e. $\overline{x^2} = u_0^2 t^2$, in agreement with physical intuition.

Thus far, we have rather naïvely used the term "average" without specifying what is exactly meant thereby. Broadly speaking, two main types of average may be recognized. On the one hand, we may focus attention on one system (e.g. an electrical resonant circuit) and observe the progress of some parameter (say current) in time from some specified initial conditions (e.g. q_0, i_0) and then form a time-average of the parameter (e.g. $\overline{i^2}$) which will, in the general case, be a function of these initial conditions. Alternatively, we may consider a large number of similar systems (e.g. collection of electrical resonant circuits with identical components) and form an average over the observed corresponding parameter for each of these systems at the same movement of time. In sufficiently large congregations which have reached statistical equilibrium these two modes of averaging will provide the same result although it may be that one may prove easier to evaluate than the other. On the other hand, if we consider equation (12) where the initial conditions are still "remembered", the value of $\overline{u^2}$ will depend on which type of average we adopt. Let us assume that we have started with a collection of particles which, as a congregation, have already achieved

overall statistical equilibrium. We now choose some individual particle, having velocity u_0 at $t=0$, and follow its progress during the subsequent time t. Equation (12) then describes essentially how the influence of the initial conditions for this particular particle "wears off" until ultimately the expected value of u^2 is simply $\overline{u^2}=\boldsymbol{k}T/m$. On the other hand, if we consider an average taken over the whole assembly of particles at any time t we should anticipate that $\frac{1}{2}m\overline{u^2}=\frac{1}{2}\boldsymbol{k}T$, since we have postulated that the congregation is initially in equilibrium. Hence in this case we interpret u_0^2 as the initial value of u^2 averaged over the assembly and we must have $\overline{u^2}=u_0^2$ and therefore

$$\overline{u^2}(1-e^{-2\beta t})=\frac{\Theta}{2\beta}(1-e^{-2\beta t}),$$

i.e. $\overline{u^2}=\boldsymbol{k}T/m$, independent of t as we expect. This question is discussed in greater detail, for example, by Ornstein (1919) in a classic paper "On the Brownian Motion."

An example where care in averaging would be required occurs if we consider the familiar sight of a queue in a restaurant. We assume that people are arriving randomly at the "tail" of the queue at some average rate and are similarly being seated, served and thus disposed of in a random manner, but at the same average rate. We then ask what may be said of the fluctuations in the length of the queue? In this case, the answer depends on whether we average with respect to one particular queue over a period of time, or take an average over a large number of queues at similar restaurants at some moment.

If it is possible to represent some random process by a (random) series of similar "shocks" each individually producing a response $f(t)$ in the observing instrument for a stimulus occurring at $t=0$, then two theorems due to Campbell (1909, 1910) are of considerable value. Let these "shocks" occur at an average rate N per unit time and let the overall instrument response be $\theta(t)$. Then it is relatively straightforward to show that

$$\bar{\theta}=N\int_0^\infty f(t)\,dt, \qquad \ldots\ldots(17a)$$

$$\overline{\delta\theta^2}\equiv\overline{(\theta-\bar{\theta})^2}=N\int_0^\infty f(t)^2\,dt. \qquad \ldots\ldots(17b)$$

Thus if, for example, particles are traversing a 100%-efficient counter randomly and each, independently of any other, produces a voltage $Ee^{-(t-t_0)/\tau}$ on arrival at $t=t_0$, the output voltage and its deviation will be given by

$$\overline{V}=N\int_0^\infty Ee^{-t/\tau}\,dt=NE\tau,$$

$$\overline{\delta V^2}\equiv\overline{(V-\overline{V})^2}=N\int_0^\infty E^2e^{-2t/\tau}\,dt=\tfrac{1}{2}NE^2\tau.$$

The theorems find ready application in thermionic and photoelectric emission, each electron transit constituting an event, or on the other hand, to such a problem as the electrical noise received from the sun if this may be regarded as originating from a random series of similar "eruptions".*

* The application to this problem was communicated privately to the writer by the Cambridge Radio Group in discussion.

These theorems may be generalized (e.g. Rowland 1936, 1937) to allow for a possible statistical distribution in the type of event resulting from an individual stimulus. Domb (*vide infra*) and Rivlin (1945) have also published extensions of the theorems to cover higher order moments (e.g. $\overline{(\theta-\bar{\theta})^3}$).

A very general division of interest can be made in the field of fluctuations analysis: on the one hand, in a particular problem one may only be concerned to determine the overall fluctuation magnitude as, for example, when we say that the overall uncertainty, $\Delta\phi$, in reading a galvanometer is given by $\frac{1}{2}\mu\overline{\Delta\phi^2}=\frac{1}{2}\boldsymbol{k}T$ where μ is the torsional constant. On the other hand, the absolute magnitude may interest us little but the character of the time-variation or "shape" may be of prime importance. These two problems meet when we enquire how the observed magnitude varies if our measuring instrument has—as always in practice to one degree or another—limitations as regards its speed or linearity of response. We find then an intimate connection between the time-variation of the fluctuation, or "correlation function" and the overall frequency spectrum of the instrument. The latter expression must often of course be interpreted as including the actual fluctuating element itself, as in the case of the torsion fibre and mirror of a sensitive galvanometer. Thus if we have a simple electrical circuit of a resistance and condenser (C) in parallel we may say immediately by the equipartition principle that $\frac{1}{2}C\overline{\delta v^2}=\frac{1}{2}\boldsymbol{k}T$, i.e. $\overline{\delta v^2}=\boldsymbol{k}T/C$ where v is the voltage across the isolated circuit. However, to observe this magnitude of fluctuation it will in general be necessary to provide a sensitive electrical amplifier and this will only have a limited frequency response. How then will $\overline{\delta V^2}$, the fluctuation of the instrument output voltage, be related to $\overline{\delta v^2}$? A similar problem occurred when Schottky (1918) first analysed the "shot" fluctuation (§ 3.2) in the current through a thermionic valve arising from the discreteness of the electron charge. If we assume the electrons to travel from cathode to anode independently of one another then in an interval of observation, τ*, the mean-square fluctuation in number crossing is given by $\overline{\delta n_\tau^2}=N\tau$, where N is the average number per second. Thus

$$\overline{\delta i_\tau^2}=\frac{\boldsymbol{e}^2}{\tau^2}\overline{\delta n_\tau^2}=\frac{N\boldsymbol{e}^2}{\tau}=\frac{\boldsymbol{e}I}{\tau}, \qquad \text{.....(18)}$$

where i_τ is the mean current over τ and I is the overall mean current. This expression is closely related in form to Einstein's expression

$$\overline{\delta q_\tau^2}=2\boldsymbol{k}T\tau/R \text{ (equation (6)).}$$

In neither case, however, is it immediately clear how an amplifier of specified frequency response will respond to noise of this character.

In the past there has been no little discussion whether noise has a frequency spectrum (cf. Fry 1925, Campbell and Francis 1946, Woodward 1948a). It is true to say that no such thing exists for the actual noise "wave" in the way that we may refer, for example, to the three-line spectrum of a simple amplitude modulated carrier. This is simply a consequence of the fact that noise itself *is* random; if such a spectrum could be defined either as an average over time or over various samples of the fluctuation from similar instruments then we could

* We shall now use τ as the interval of observation, as we wish to reserve t as an arbitrary time-variable.

predict, by Fourier inversion, how the noise variable would proceed in time. In fact, the average noise-spectrum (in either case) is zero, which means essentially that if we try to predict therefrom the future course of the fluctuation we simply obtain zero. This in turn means that the average, or "expected" noise amplitude at any given time t, is zero *—an entirely correct result, but not one of any great value! If, however, we enquire how the mean square noise (or noise "power") is distributed in frequency this leads, in the limit, to a perfectly definite conclusion. This is closely related to the fact that if we consider the noise at any instant t, and its value at time τ later, then these values are *not* in general random with respect to one another; if, for example, we take their product and average over time or equivalent systems we achieve again a quite definite result, characteristic of the type of fluctuation, which is a function of the correlation interval τ. Noise from a lightly damped resonant system, for example, will have a quasi-periodic character, the amplitude and phase of the oscillation varying irregularly (see Figures 1 and 2). A little thought will show that the correlation product too

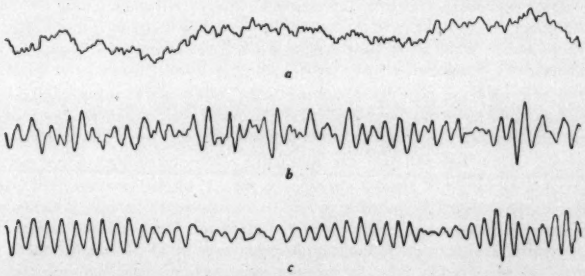


Figure 1. Record of Brownian movement of sensitive torsion balance. (*After* Kappler) (*a*) Atmospheric pressure. (*b*) Pressure 1×10^{-3} mm. Hg. (*c*) Pressure 1×10^{-4} mm. Hg.

will exhibit a quasi-periodic character. For small values of τ the cyclic correlation will be almost perfect and the correlation function will oscillate regularly, but for large τ, the periodic character will be smeared out as the degree of correlation between the two values, at t and $t+\tau$, diminishes. The relationship may usefully be compared with that of a diffraction pattern to the grating producing it. More complex correlation functions may be devised involving, for example, the time-derivatives of the fluctuation. Such expressions will in fact increase our knowledge of the precise character of the noise, but for very many purposes it is sufficient to know either the standard correlation function or its complement, the (power) frequency-spectrum.

The analytic expression of the foregoing is known as the Wiener–Khintchine theorem (Wiener 1930, Khintchine 1934). We define the correlation function $\psi(\tau)$ of the random variable $y(t)$ by $\psi(\tau)=\overline{y(t)y(t+\tau)}$, where the bar denotes averaging, which may generally speaking be with respect to time for a single

* We neglect a possible "bias" or "D.C. component"; the removal of this is a trivial requirement.

system or over a congregation. Then the distribution in frequency $w(f)$ is given by

$$w(f) = 4\int_0^\infty \psi(\tau)\cos 2\pi f\tau\, d\tau, \qquad \ldots\ldots(19a)$$

(where $\overline{y^2} = \int_0^\infty w(f)\,df$) and by Fourier inversion we have

$$\psi(\tau) = \int_0^\infty w(f)\cos 2\pi f\tau\, df. \qquad \ldots\ldots(19b)$$

If then the fluctuation is presented to an amplifier whose power spectrum is $A(f)$ the observed output of mean square fluctuation $\overline{\delta Y^2}$ is

$$\overline{\delta Y^2} = \int_0^\infty A(f)w(f)\,df.$$

If the correlation interval—that is, roughly the "relaxation time" of the system—is too short for our measurement methods to resolve then the correlation function is practically zero everywhere except for $\tau \to 0$ (i.e. a Dirac "impulse-function") and the power frequency-spectrum is uniform. This is the case for the idealized mathematical concept of the track of a Brownian particle, which is frequently referred to as "non-differentiable" (cf. Bouligand 1934).

Many workers have used and added to the knowledge of the value of the correlation function. We mention particularly the work of Uhlenbeck and his colleagues (Uhlenbeck and Ornstein 1930, Wang and Uhlenbeck 1945), Burgess of the National Physical Laboratory (unpublished), and Rice (1944, 1945, 1948) of Bell Telephone Laboratories. The recent papers of the latter worker form a monograph dealing with the analysis of a wide variety of electrical noise problems although of course the underlying methods are universally applicable. Milatz and Van der Velden (1943) have analysed the frequency-distribution of temperature fluctuations starting from the appropriate Langevin-type equation to evaluate the correlation function. This problem, also discussed independently by Daunt and Kompfner (1945–1946), arises in the analysis of the limiting sensitivity of thermal detectors. The method of analysis and general conclusions are of course equally applicable to electrical and mechanical systems with suitable interpretation of the parameters involved.

Consider a body of thermal capacity C connected to its surroundings by a thermal resistance R. We imagine the surroundings as a "heat-bath" and denote the temperature difference of C by θ. Then the appropriate Langevin equation is

$$C\frac{d\theta}{dt} + \frac{1}{R}\theta = H(t), \qquad \ldots\ldots(20)$$

where $H(t)$ represents a random spontaneous heat flow with properties similar to those of $F(t)$ in equation (10).

Then

$$\theta(t) = \theta_0 e^{-\beta t} + e^{-\beta t}\int_0^t e^{\beta\xi}h(\xi)\,d\xi$$

with $\beta = 1/RC$, $h(t) = H(t)/C$;

$$\theta(t+\tau) = \theta_0 e^{-\beta(t+\tau)} + e^{-\beta(t+\tau)}\int_0^{t+\tau} e^{\beta\zeta}h(\zeta)\,d\zeta.$$

Then

$$\overline{\theta(t)\theta(t+\tau)} = \theta_0^2 e^{-\beta(2t+\tau)} + e^{-\beta(2t+\tau)} \overline{\int_0^t \int_0^{t+\tau} e^{\beta(\xi+\zeta)} h(\xi) h(\zeta)\, d\zeta\, d\xi},$$

leading to

$$\overline{\theta(t)\theta(t+\tau)} = \theta_0^2 e^{-\beta\tau} e^{-2\beta t} + \frac{e^{-\beta\tau}}{2\beta}\{1 - e^{-2\beta t}\} \,.\, \Psi,$$

where $\Psi = \int_{-\infty}^{\infty} \overline{h(\xi)\, h(\xi+\nu)}\, d\nu$, bearing in mind that $\overline{h(t)\, h(t+\nu)} \simeq 0$, except for $\nu \simeq 0$. Thus in equilibrium ($t \to \infty$): $\psi(\tau) \equiv \overline{\theta(t)\theta(t+\tau)} = (e^{-\beta\tau}/2\beta)\Psi$. We know, however, from general considerations (cf. $\frac{1}{2}mu^2 = \frac{1}{2}\boldsymbol{k}T$ in the mechanical case) that $\overline{\theta(t)^2} = \boldsymbol{k}T^2/C$. Therefore $\psi(0) \equiv \overline{\theta(t)^2} = \Psi/2\beta = \boldsymbol{k}T^2/C$. Hence $\psi(\tau) = (\boldsymbol{k}T^2/C)e^{-\beta\tau}$. Thus

$$w(f) = \frac{4\boldsymbol{k}T^2}{C} \int_0^{\infty} e^{-\beta\tau} \cos 2\pi f\tau\, d\tau \quad \text{(from equation (19}a\text{))}$$

$$= \frac{4\boldsymbol{k}T^2}{C} \frac{\beta}{\beta^2 + 4\pi^2 f^2} = \frac{4R\boldsymbol{k}T^2}{1 + \omega^2 C^2 R^2} \quad (\omega = 2\pi f).$$

A class of problem of wide interest, experimentally and theoretically, is that of noise whose frequency spectrum is more or less confined within relatively narrow limits about some frequency f_0—corresponding in the electrical case to noise from a " narrow band-pass " filter. The fluctuation is then expressible as

$$y(t) = R(t) \cos\{\omega_0 t + \phi(t)\}, \qquad \text{.....(21)}$$

where the " envelope " or modulus R and the phase ϕ are random functions varying relatively slowly in comparison with $\cos \omega_0 t$. The probability distribution of R follows the Rayleigh law:

$$p(R)dR = (2R/\overline{R^2}) \exp(-R^2/\overline{R^2})\, dR \qquad \text{.....(22)}$$

and that of ϕ simply $p(\phi)d\phi = (1/2\pi)\, d\phi$.

Much more interest, however, attaches to such compound probability functions as $p(R_1(t), R_2(t+\tau))\, dR_1 dR_2$ which determine the correlation and hence the " character " of the noise. Say, for example, as Burgess and Uhlenbeck have done, we wish to determine the character of such a fluctuation after passage through a non-linear device (" detector ") having a response $E = kR^n$. Starting from a knowledge of $p(R_1, R_2)$ we can then obtain an expression for the correlation of the output:

$$\overline{E_1(t)E_2(t+\tau)} = \overline{k^2 R_1^n(t) R_2^n(t+\tau)},$$

whence, by application of (19a), the form of the output spectrum is obtained. This may in fact differ markedly from the spectrum before entry to the detector—a feature not always fully appreciated in earlier experimental work on the subject. This systematic method of approach to problems of this type scores in clarity and directness over earlier work (cf. Ragazzini 1942) in which the more technical concept of cross-modulation was used in the analysis.

An interesting feature of Rice's work (Bennett and Rice 1934, Rice 1944, 1945) is the use of contour integrals to represent the response of a non-linear device.

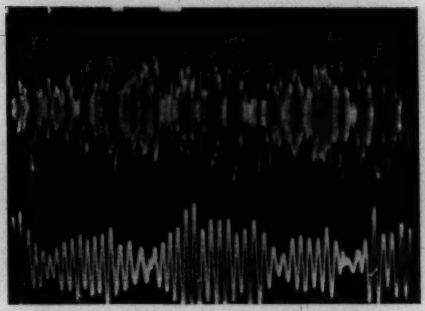


Figure 2. Upper trace: record of spontaneous electrical fluctuation. Lower trace: after rectification and passage through low-frequency (1 kc/s.) filter of narrow response.

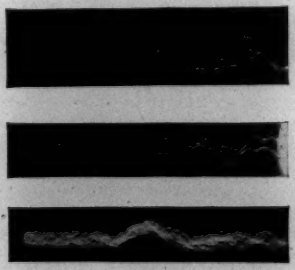


Figure 5. Photographs of single and multiple traces of pulse signal and noise: (*a*) single trace, (*b*) double trace, (*c*) forty traces.

[Conditions: Pulse duration, 1 μsec. I.F. Signal/noise power ratio, 2/1. I.F. Bandwidth, 5 Mc/s. Video bandwidth, 2 Mc/s.]

[After Bradfield *et al.*]

To face page 64

Thus a detector of response

$$E = a_1 v + a_2 v^2 + \ldots + a_n v^n; \quad (v>0)$$
$$= 0 \qquad\qquad\qquad\qquad ; \quad (v<0)$$

may be represented by

$$E = \frac{1}{2\pi i}\int_C \frac{e^{izt}}{z} \sum_{k=1}^{n} \frac{k!\, a_k}{(iz)^k}\, dz,$$

where C is a contour covering the real axis and indented downwards around the origin. The final contour integral arising at the end of the analysis may then be interpreted by the use of the inversion theorems of operational calculus. Kurt Fränz (1940, 1941, 1942) in his papers during the war in the same field of statistical analysis has also used this method. An important practical application of analysis of this type lies in the determination of the modification of effective signal/noise ratio in a Radar receiver on passage through a detector for presentation on a cathode-ray tube display.

Statistical work of this kind has proved of value in the interpretation of experimental studies in the diffraction of radio waves by the ionosphere (Ratcliffe and Pawsey 1933, Pawsey 1935, Ratcliffe 1948). The Cavendish Laboratory Radio Group under Ratcliffe have made extensive observations on the character of signals "reflected" from the ionosphere; continuous fluctuation ("fading") is observed and this has led to the proposal that the reflecting medium may be regarded as composed of many scattering elements moving with, say, a Gaussian velocity distribution. Since each scatterer will cause a frequency-change dependent on its velocity due to Doppler effect, the returning wave will be made up of a quasi-continuum of small, randomly-related, oscillatory components with a frequency-spread about either side of the upgoing frequency. This should then behave as a "noise-wave" of the type above (equation (21)); its correlation function experimentally determined enables deductions to be made about the frequency-spread involved (hence the mean velocity of the scatterers) and other parameters of interest such as the dependency of fading-rate on transmitter frequency. In this work Ratcliffe (1948) has found useful the studies by Fürth and the writer (1946, 1947a) in electrical noise of this type, and more recently (MacDonald 1949a) it has also been found of value to develop an analysis of the correlation of the phase, $\phi(t)$, which can then confirm and amplify the deductions from observations of envelope-correlation.

A rather novel illustration of the correlation concept appears in a recent paper by R. Wilson (1947) entitled "Random selectors for extra-sensory-perception experiments". In this field a requirement exists for devices which in a series will yield at each "trial" an entirely random choice from, say, six types of event. It is straightforward to ensure that the first trial will provide equal *a priori* probabilities but considerable care must be exercised in the instrument design to ensure that successive trials at say $\frac{1}{2}$-second intervals will exhibit no appreciable intercorrelation.

Bloembergen, Purcell and Pound (1948) have analysed the effect of the magnetic field fluctuations encountered by nuclei in investigations on nuclear relaxation phenomena. They employed the correlation function whence the frequency-dependence is then derived, which enables the line-width of the resonance and the relaxation-time to be deduced.

Finally, we consider one or two general deductions about the nature of the correlation function. In general, the correlation function $\psi(\tau)$ will have a maximum at $\tau=0$ and a form somewhat as in Figure 3. If $\tau \simeq 0$ then $\psi(\tau) \simeq \overline{y^2}$, a positive definite quantity; as τ increases, $y(t)$ and $y(t+\tau)$ become less correlated and their product at any given moment will vary in sign reducing the mean value of the product, $\psi(\tau)$. When $\tau \to \infty$, $y(t)$ and $y(t+\tau)$ become entirely uncorrelated so that $\psi(\tau) \to 0$.* Consider now the correlation between $y(t)$ and its subsequent time-derivative or "velocity" $\dot{y}(t+\tau)$. Then one shows immediately that

$$\overline{y(t)\dot{y}(t+\tau)} \equiv \phi(\tau) = \frac{\partial}{\partial \tau}\psi(\tau),$$

and therefore has the form in Figure 4. Thus $\overline{y(t)\dot{y}(t)} = 0$; that is, at any arbitrary instant the function and its velocity are uncorrelated. This merely signifies that at any moment the function is as likely to be in the act of increasing or decreasing—a property of its essentially random nature. On the other hand, however, $\overline{y(t)\dot{y}(t+\tau)} < 0$ for $\tau > 0$. This then demonstrates a very significant property; if at some instant the function has a positive deviation from zero it is then at some time τ later more likely to be diminishing than increasing. Again, of course, if τ becomes very large the correlation between $y(t)$ and $\dot{y}(t+\tau)$ vanishes (i.e. $\phi(\tau) \to 0$ as $\tau \to \infty$). One appreciates thus the behaviour and significance of spontaneous fluctuations in the field, for example, of thermodynamics in the tendency "almost always" to be returning towards the equilibrium position. This characteristic is also exhibited by observing that $\overline{y(t)\ddot{y}(t)}$ $\left(= \left[\frac{\partial^2}{\partial \tau^2}\psi(\tau)\right]_{\tau=0}\right)$ is essentially negative; that is, on the average, the system is being "attracted back" towards equilibrium.

Another pair of inversion theorems relating the temporal behaviour of noise to its frequency distribution has been noticed recently by the writer (MacDonald 1949b). If $\overline{y_\tau^2}$ be the mean square value of the fluctuation taken over any interval τ, and $w(f)$ the power-spectrum as in (19a), then we find:

$$\Psi''(\tau) = 2\int_0^\infty \frac{w(f)}{2\pi f}\sin 2\pi f\tau \,.\, df, \qquad \ldots\ldots(23a)$$

$$w(f) = 4\pi f \int_0^\infty \Psi''(\tau)\sin 2\pi f\tau \,.\, d\tau, \qquad \ldots\ldots(23b)$$

where $\Psi''(\tau) \equiv \frac{\partial}{\partial \tau}\{\tau^2 \overline{y_\tau^2}\}$.

If we take Einstein's expression (6) for thermal fluctuation of electric charge then it is evident that

$$\overline{\delta i_\tau^2} = 2kT/R\tau, \qquad \ldots\ldots(24)$$

where δi is the effective current fluctuation. From this it now follows that the current fluctuation is effectively distributed uniformly in the power-spectrum

$$w(f)\,df \equiv \overline{\delta i_f^2} = \frac{4kT}{R}\,df,$$

the well-known result deduced by Nyquist by a fresh thermodynamical argument (1927, 1928). Similarly, the much more familiar expression for the shot effect,

* We assume, as before, that $\overline{y(t)} = 0$.

$\overline{\delta i_f^2} = 2e\bar{I}df$, follows directly from Schottky's original equation (18). The writer has used the theorems to derive the progressive disordering—due to random thermal velocities—of an electron beam, initially ordered in density, while drifting in time (e.g. in a klystron).

In many processes the events obey the Poisson or Gauss distribution law; it is then relatively straightforward to calculate the fluctuations. This, on the other hand, is not so in such cases as the number of particles produced in a cosmic-ray shower nor in chain or branch radioactive decay. Ellickson (1938) has used a generating function for calculating probabilities and fluctuation parameters in this type of problem and Woodward (1948b) more recently in the theory of cascade multiplication. Rice (1944, 1945), too, has used this device in his work. We consider here a simple case obeying in fact Poisson's law to illustrate the method. Events (e.g. emission of electrons from a cathode) are occurring purely randomly

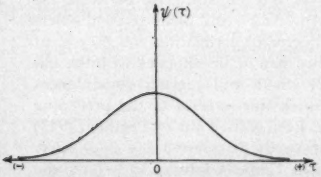


Figure 3. Sketch of typical correlation function $\psi(\tau)$ of a random variable.

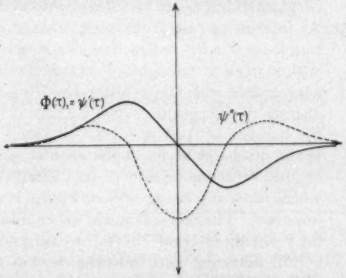


Figure 4. Sketch of derivatives of typical correlation function of a random variable.

at an average rate λ per unit time such that the probability of an event occurring within the infinitesimal time-interval dt is λdt, this being independent of the previous history or subsequent behaviour of the system. Let the probability of n events occurring within the time t be $P(n, t)$. Then it is easy to show that

$$(d/dt)P(n, t) = \lambda\{-P(n, t) + P(n-1, t)\}$$

subject to the initial conditions $P(0, 0) = 1$; $P(n, 0) = 0$ $(n \neq 0)$. This equation may in fact be solved by successive integration (cf. Furry 1937). Let us rather, however, define a generating function $Q(x, t)$ such that $Q(x, t) = \sum_{n=0}^{\infty} x^n P(n, t)$ (i.e. $P(n, t)$ is then the coefficient of x^n in the expansion of Q in powers of x).

We can then show that Q satisfies the differential equation *

$$\frac{\partial}{\partial t} Q(x, t) = \lambda(x-1)Q(x, t), \text{ with the initial condition } Q(x, 0) = 1.$$

Hence
$$Q(x, t) = e^{\lambda(x-1)t}. \qquad \ldots\ldots (25)$$

Expanding $Q(x, t)$ in powers of x we find then that $P(n, t) = e^{-\lambda t}\dfrac{(\lambda t)^n}{n!}$, the familira

* In this simple example Q satisfies an ordinary differential equation in t; more generally however, a partial differential equation in t and x will emerge.

Poisson law. However, if we only wish to find, say, the mean value $\bar{n}$ and the mean square deviation $\overline{\delta^2} \equiv \overline{n^2} - (\bar{n})^2$, as often occurs, this may be deduced directly from a knowledge of $Q(x, t)$ alone.

Thus
$$\bar{n} = \sum_{n=0}^{\infty} nP(n, t) = \left[\frac{\partial}{\partial x} Q(x, t)\right]_{x=1},$$

which, in this case $= [\lambda t e^{\lambda(x-1)t}]_{x=1} = \lambda t$, using (25).

And
$$\overline{n^2} = \sum_{n=0}^{\infty} n^2 P(n, t) = \left[\frac{\partial}{\partial x} x \frac{\partial}{\partial x} Q(x, t)\right]_{x=1}$$
$$= \lambda^2 t^2 + \lambda t, \text{ here,}$$

giving $\overline{\delta^2} = \lambda t$, a well-known result.

The calculation of random probability functions, although not necessarily of fluctuations, arises very frequently in the analysis of particle-counter behaviour. An interesting case arises when, instead of measuring the entirely random radiation from a pure radioactive source, we examine for example a source producing within itself a "daughter" radioactive element, if the latter has a half-lifetime comparable with the counter resolving-time or the mean interval between discharges of the counter. Such a case is the An→AcA→AcB disintegration where the half-life of AcA is $\sim 1{\cdot}8 \times 10^{-3}$ sec. A degree of correlation will then exist in the discharge series of the counter, and this may be investigated with the use of the "interval-selector" (cf. Ward 1942) which will register coincidences whenever a discharge occurs within a specified time-interval of the preceding impulse. The relevant statistical analysis has been carried out by Feather (1943) for "pairing-interval" both constant and statistically distributed.

All detecting and counting devices such as Geiger counters and electronic scaling circuits have a characteristic limit of resolution due to the finite time required for return to the equilibrium state after registration of the event. The conventional theory (e.g. Volz 1935, Strong 1938) is usually based on the assumption of a single constant time, τ, which must elapse after an event—whether successfully recorded or not—before the instrument can record a subsequent event. Klemens (1948) has generalized the statistical theory to permit inclusion of two recovery times, τ_1 and τ_2, where τ_1 is the time necessary after a *successful* registration before another event can be recorded, while τ_2 is the time required after an event which was *not* recorded (i.e. due to saturation of the counter by some previous event) before another event can be successfully recorded.

One of the most classic problems in probability is the analysis of the distribution of the modulus of the sum of a large number, n, of independent vibrations of random phase, i.e. vectors of the type $a_k \exp i\phi_k$. Rayleigh (*Theory of Sound*, vol. 1, § 42a) considered the cases where all $a_k = \alpha$, say, and also where the individual amplitudes a_k are distributed in a number of large groups. Domb (1946) has extended the analysis to cover a dependence of vector amplitude, a_k, on the individual phase, ϕ_k, and has applied the method used to the shot effect. Miss M. Slack (1946) has encountered a specialized Rayleigh problem in the field of multi-channel communication and has derived the distribution functions of amplitude for *small* numbers ($n = 1, 2, 3, \ldots$.) of independent vibrations. The Rayleigh distribution is met with in the classical "random-walk" problem (Pearson 1905a, b, Rayleigh 1905), and Miss Slack's results correspond in this case to the initial few steps before the "walk" has degenerated into an essentially random process.

§3. SURVEY OF DEVELOPMENT IN RECENT YEARS

3.1. *Electrical Noise*

The demands of the recent World War in the field of Radio produced very considerable development of Radar and Communications technique. In Radar, specifically, one has to detect and display a signal reflected, or re-radiated, from a target subtending a rather small solid angle; it is then easy to see that the problem of maintaining the signal, during the high degree of amplification necessary to provide visibility on, say, a cathode-ray tube, above the level of the inherent electrical noise is a very severe one. Consider a transmitter at ground-level emitting a signal-power P_T from an aerial of power-gain G_T—free from the complications of ground reflection—the signal being reflected from a target of effective reflecting area σ at range R to a receiving aerial of effective absorptive area A_R. We may readily show (e.g. Norton and Omberg 1947) that the received power P_R is then

$$P_R = \frac{G_T P_T \sigma A_R}{16\pi^2 R^4}. \qquad \ldots\ldots(26)$$

If then the overall effective noise power generated in the receiver, referred back to the input terminals, be P_N, a target may be expected to be detectable so long as $P_R \gtrsim P_N$. The maximum range is thus

$$R_{max} \simeq \sqrt[4]{\frac{G_T P_T \sigma A_R}{16\pi^2 P_N}}. \qquad \ldots\ldots(27)$$

To double the maximum range, therefore, it will be necessary to increase the transmitter power by $2^4 = 16$.* Since it is evident from (27) that an exactly parallel improvement can be provided by reduction of P_N in the same ratio, research into methods of improvement of signal/noise ratio P_R/P_N was of great importance. We remark further, that given a certain ratio P_R/P_N, the corresponding *observable* signal/noise ratio will depend on the type of display employed. Considerable effort has gone towards determination of optimum presentation conditions (e.g. Hopkinson 1946). A vital factor is the recurrence frequency of pulse transmission which determines effectively the number of opportunities presented to the operator in a given time for contrasting the radar signal and the background noise; this compares with the improvement in relative counting accuracy in, say, a Geiger counter with increase in significant counting-rate, or the increase in effective sensitivity of a galvanometer against Brownian movement with increase in observation period (Ornstein and Milatz 1938). The importance of this factor is well depicted in the cathode-ray photographs (see Figure 5) due to Bradfield *et al.* (1946). A general deduction of interest is that the conventional "A-type" presentation (as in Figure 5) proves superior to other methods, including meter-displays, in the detection of limiting signals through noise.

In assessing the influence of various factors, such as degree of aerial coupling and amplification of individual stages, a need was evident for some standard specification of signal/noise performance (Strutt and v. d. Ziel 1941, 1946, North

* We may remark further that where ground reflection is a prime factor $R_{max} \sim (P_T)^{\frac{1}{4}}$ and thus doubling of R_{max} would now demand an increase in P_T of $2^8 : 1 = 256 : 1$ pointing the moral even more strongly!

1942, Friis 1944, MacDonald 1944, v. d. Ziel 1947a). This led to the definition of noise figure or noise factor, F, as follows. Let the (maximum) available signal- and noise-power from the aerial terminals be p_s and p_n respectively* and P_{so} and P_{no} the available signal and noise power from the output terminals under operating conditions. That is to say, the receiver may not necessarily in general be coupled to the aerial terminals in such a manner as to absorb p_s and p_n, since this condition does not as a rule coincide with optimum signal/noise performance. Then

$$F=\frac{P_{no}}{P_{so}}\Big/\frac{p_n}{p_s}. \qquad \ldots\ldots(28)$$

In general, $F>1$ for any physical receiver. A number of workers (e.g. Strutt and v. d. Ziel 1942a, b, Burgess 1943, 1945, 1946a) have considered general requirements for achieving optimum noise-figure including the effects of feedback where applied. We should also mention here the systematic method of noise analysis in complex valve circuits developed by the team of Campbell, Francis and James (1945, 1946).

Returning to equation (27) we use $G_T=\dfrac{4\pi A_T}{\lambda^2}$ where A_T is the effective radiation area. Thus

$$R_{max}\simeq\sqrt[4]{\frac{P_T A_T A_R \sigma}{4\pi P_N}}\cdot\frac{1}{\sqrt{\lambda}}. \qquad \ldots\ldots(29)$$

This expression provides a compelling reason for progress towards very short wavelengths; for a given physical size of equipment (A_T, A_R), which may tactically be of prime importance, increased range can be rapidly obtained by reduction of wavelength. But of course such improvement depends on maintenance of the noise power, P_N, at its previous level, and consequently at very short wavelengths there was continual effort to prevent deterioration of the signal/noise ratio. The problem may be readily appreciated by considering the composition of the noise-factor. At the higher frequencies it becomes increasingly difficult to amplify a signal and the obvious solution is to convert ("mix") to a lower frequency immediately. If then the mixer-unit has a noise-factor F_1 and power-gain G_1 and the intermediate-frequency amplifier noise-factor F_2 then

$$F=F_1+(F_2-1)/G_1. \qquad \ldots\ldots(30)$$

At the highest frequencies a diode or crystal-valve has to be employed (with, of course, a local oscillator) as frequency-changer; hence $G_1<1$ necessarily. Although the mixer itself may not necessarily be *inherently* particularly "noisy" (in fact a common misconception) the reduction in G_1 will throw into relief the noise of the main amplifier against the attenuated incoming signal.† Thus even in centimetric receivers the improvement of noise-factor in I.F. amplifiers was a primary concern: at the same time of course very considerable research and development went into the production of crystals (e.g. Bleaney *et al.* 1946a, b). The actual mechanism of noise generation in a crystal is as yet not clear. At the origin of the characteristic—when no power is being delivered from the local

* p_n is conventionally taken equal to $kT\Delta f$, where T is ambient temperature.

† For an average production crystal the noise temperature (i.e. measuring the excess noise over thermal noise) is as low as 1·6, while the conversion gain is only 0·2.

oscillator—it obeys Nyquist's (thermal-noise) formula in terms of its slope-resistance. When, however, the crystal is under operation power is being injected from the oscillator; a consequent increase in noise-temperature over that obtaining in thermal equilibrium is certainly not unexpected. Detailed explanation may be sought in terms of a shot-effect mechanism (e.g. Bleaney *et al.* 1946b); another contributory factor may be the existence of (thermally) unstable micro-contacts between the silicon and tungsten "cat's-whisker" comparable perhaps with "crazy-contacts" in certain semiconductors (cf. Christensen and Pearson 1936). The problem of noise in semiconductors (see § 3.3) is still under investigation; we may hope for further clarification.

It should then be evident that if even a relatively small degree of power-amplification could be provided in a reasonably low-noise stage before the first detector the situation would greatly improve. Success was obtained up to frequencies of approximately 600 Mc/s. through introduction and development of the grid-return triode (grounded-grid triode) (Bell 1946, Foster 1946). A fundamental feature is the reduction of interaction between output and input circuits via inter-electrode capacity.

At the highest frequencies the travelling-wave amplifier (Kompfner 1947 a, b, c, Pierce 1947) appears to offer excellent promise of reasonable amplification without introduction of excessive noise. Other types of centimetre-wave amplifier, such as the klystron (very familiar as an oscillator), suffer essentially from the disadvantage that direct interaction of the electron beam with the input rhumbatron introduces severe noise into that circuit with disastrous results on the signal/noise performance. Strutt and v. d. Ziel (1948), however, in a paper entitled "Application of velocity-modulation tubes for reception at U.H.F.* and S.H.F.†" have said that ". . . . the application of velocity modulation pre-amplifiers is not an entirely hopeless project if special noise-reducing measures are taken into account. These measures consist of the introduction of a pre-circuit between the electron gun and the input circuit." Kompfner (1947b) has also pointed out that the shot noise directly induced in the "catcher" (output rhumbatron) by the random electron beam has the same essential origin as the "amplified shot-noise" appearing in the output; the latter arises from modulation of the electron beam by fluctuation voltages induced directly in the "buncher" (input rhumbatron). The primary origin of both effects is, of course, the fundamental shot fluctuation of the beam and, therefore, as Kompfner says: ". . . . it should be possible, in principle, to play one against the other with . . . . partial compensation."

Nevertheless, since the noise figures of conventional velocity-modulation tubes amount to a few *thousand* (Strutt and v. d. Ziel 1948) the possibility of reducing this at all readily to compare with a crystal mixer receiver ($F \sim 10$) seems rather remote.

The travelling-wave tube avoids these difficulties effectively inasmuch as the amplification process is spread out over the whole tube-length; the coupling of the electron beam to the input circuit is therefore less important than in a klystron. An alternative view-point is that the given electron beam is used very much more efficiently for amplification; a specified degree of amplification may thus be achieved for a much lower beam-current and correspondingly lower noise.

* Ultra-high frequencies. † Super-high frequencies.

The question of noise reduction in klystrons has its analogue in the valve of "normal" design in "induced grid noise" discovered independently by Bakker (1941) and North and Ferris (1941). Density fluctuations in the anode current stream induce significant current fluctuations in the grid circuit when the transit-time becomes appreciable. These current fluctuations will develop a grid-cathode voltage across the effective input circuit impedance and hence modulate the anode current. It is again evident that these current variations will be directly phase-related to the primary shot fluctuations and hence suitable detuning of the input circuit from resonance towards capacitive reactance can provide partial suppression. This correlation was recognized by Strutt and v. d. Ziel, experimental confirmation being furnished by Kleen (1941) and v. d. Ziel and Versnel (1948), who observed at 7·25 m. wavelength considerable asymmetry in noise resonance curves due to the interaction of "shot" and "induced grid noise", which was not found when a large noise excitation was injected from an external source. A closely-related phenomenon is that of "total-emission noise", first predicted by Smyth (1946). When transit-time becomes appreciable, those electrons which are emitted from the cathode and return thereto after repulsion from the space-charge barrier can provide a component of input loss-conductance (v. d. Ziel 1947 b). On very general grounds (§ 3.2), one would therefore anticipate that this would form a source of further "induced" fluctuation in the grid-cathode circuit; that this is indeed so has been experimentally verified by v. d. Ziel and Versnel (1947, 1948) again in diodes at 7·25 m. wavelength and the effective noise temperature is found to be practically that of the cathode. Since the group of electrons *not* crossing the space-charge barrier is in almost perfect thermal equilibrium with the cathode this conclusion must be regarded as entirely satisfactory.

Clearly a vital feature in such developments must be the provision of means for the reliable measurement of noise—at the high frequencies particularly. Fundamentally, two reference sources exist: thermal noise from an element obeying Nyquist's law, or shot noise from a saturated diode in accord with Schottky's equation. The latter method has been used by several workers in fundamental noise studies at moderate frequencies; with the employment of a tungsten filament diode of small electrode spacing and low lead inductance and with suitable circuit design etc., the method can be carried up to approximately 300 Mc/s. (Moffatt 1946). Transit-time damping and impedance uncertainties present stumbling-blocks at the highest frequencies. Kompfner *et al.* (1946) have developed a diode noise-source of close-spaced transmission-line construction which surmounts these difficulties; the impedance involved is then that of the line—calculable from the geometry—and the appropriate transit-time correction may also be accurately computed. Ullrich and Rogers (1946) have employed the thermal fluctuation of a resistive element at high temperature as a standard source of electrical noise. A small lamp, of bolometer type, with a resistance of approximately 70 Ω (simulating standard coaxial transmission-line impedance) is the heart of this method, which is reported to agree well with other methods up to about 600 Mc/s. Satisfactory results have not, however, been obtained in the 3,000 Mc/s. region.

Note added in proof. More recently some doubt has been cast on the published limits of accuracy etc. of these noise measurement methods, which are considered in general to be too optimistic.

3.2. *Fundamental Valve Noise Problems*

The inter-relationship between thermal and shot noise has provoked much thought in the past. As mentioned earlier, Einstein first recognized that fluctuations of electricity must arise from molecular thermal agitation—itself consequent upon the discrete, or "grainy", character of fundamental matter. These fluctuations—"thermal noise"—are, however, entirely independent of any assumptions about the actual structure of the electrical fluid; all that is essentially required is an interaction between the electricity and the thermal motion of the material molecules whereby the thermal fluctuations of the molecules may be manifested in the electricity. The situation is entirely analogous to the mechanical ("collision") interaction between the molecules of a fluid and a macroscopic Brownian particle suspended therein. The interaction is expressed through the electrical resistance, R, of a given element while the discrete nature of atomic matter appears in Boltzmann's constant, k. The intimate association between dissipation and thermal fluctuations is underlined; we may be certain that in any system where thermal fluctuations are evinced a dissipative component must be present.

Of a rather different character, however, are the irregular fluctuations in a flowing electric current which can arise from the atomicity of electricity *itself*. This type of fluctuation, the shot effect, was recognized in photoelectric and associated studies by von Schweidler (1905, 1910, 1913) and Campbell (1909, 1910) and in the field of thermionic emission in a classic paper by Schottky (1918). In the case of the thermionic valve a certain amount of confusion has arisen, essentially because the electron emission is there provided from a thermal source. When a diode—to take the simplest structure—has a sufficiently high anode potential applied (the "saturation" state) then every electron emitted from the cathode proceeds to the anode, contributing directly to the observed valve current. It is then evident that only fluctuations in the *number* of electrons emitted can play a part in the observed noise since variations of emission energy cannot affect the transit of any other emitted electron and the external circuit only responds to current (i.e. number of electrons crossing a given plane per unit time).* One would in this case only expect to encounter the quantum of charge e and the number flowing as fundamental parameters, as is in fact the case in the simple shot equation (cf. equation (18)).

When, however, a valve is operated under conditions of space-charge limitation a new state of affairs obtains. Electrons of various energies can now interact with one another to a certain extent through the medium of the space-charge. This is generally expressed by saying that the potential minimum in the valve fluctuates in a sense tending to minimize the effect of a "primary" fluctuation of the emission, the net reduction in the observed noise being formally included by a dimensionless factor $\Gamma^2(\leqslant 1)$ in the shot equation. A second distinctive feature of the valve now is the appearance of a finite differential resistance, R_a. Variation of the external potential difference can alter the magnitude of the valve current—a feature not present in the saturated valve. This control depends directly on the variability of the potential minimum; it follows that the new phenomena will be inter-related. It is now clear (e.g. North *et al.* 1940–1941, Schottky and Spenke 1937) that in the fluctuation problem the reduction effect

* This statement is of course restricted to conventional structures where transit-time is negligible.

arises from correlation of the fluctuation in the number of electrons with energies very close to that of the minimum with the fluctuations in electrons of other energy-classes. Similarly, when altering the anode potential, the variation in (the mean) current is provided solely by electrons with energies close to the minimum. The individual electron charge now ceases to be the primary factor; rather the energy distribution—characteristic of the cathode temperature—enters through the familiar factor kT, and we are not surprised to find that the anode current fluctuation is *now* expressible in a form closely analogous to that for thermal fluctuations in a passive resistance $R = R_a$. On the other hand, the effective noise temperature is not identical with the emitter temperature since those electrons composing the valve current (generally gaining energy at the expense of an external battery) are not in thermal equilibrium with the cathode.

The transition from saturation to space-charge limitation, in particular, thus appears to the writer of fundamental significance; partly in consequence of this he has found himself (MacDonald 1948a) unable to agree with Fürth's (1948b) recent general analysis of valve noise based on a treatment of the thermionic valve under arbitrary operational conditions as being effectively in a state of thermal equilibrium.

Related to the foregoing discussion is the question of current-fluctuations in a photocell (emissive) and the incident radiation. If the cell, as is customary, be operated at saturation, then it appears clear *a priori* that no information about the energy distribution of the incident photons can be derived since solely the *number* of electrons released is significant. The observable current fluctuation will be the same for a low- or high-temperature radiation source productive of the same emission current. It has sometimes been asked however (e.g. Pirenne, discussion on Fürth and MacDonald 1947a) whether information about the rate of arrival in, say, a monochromatic beam can be obtained. The answer to this depends essentially on the quantum-efficiency of the device. The problem, in its essentials, may readily be analysed (cf. MacDonald and Kompfner 1949): if p be the probability of emission per photon, I the mean cell-current, and N the mean rate (with fluctuation $\overline{\delta N^2}$) of photon-arrival, then the fluctuation in the electron current is

$$\overline{\delta I^2} = 2eI\left\{\frac{\overline{\delta N^2}}{N}p + (1-p)\right\}\Delta f.$$

If the photons arrive in simple random fashion, then $\overline{\delta N^2} = N$ * and the electron-current is purely random obeying Schottky's formula $\overline{\delta I^2} = 2eI\Delta f$, independent of p; on the other hand, if $p \ll 1$, $\overline{\delta I^2} \simeq 2eI\Delta f$ again, independent of the mode of photon-arrival. In modern emitter cells, p may rise to approximately 0·2 at the selective wavelength peak and were any significant degree of correlation to exist it should then be within the limits of experimental observation.

If, however, the cell functions under space-charge limited conditions then the emission-energy distribution can play a part by mutual interaction through the potential minimum mechanism. A preliminary investigation of the electron noise-temperature under these conditions has been reported (Fürth and MacDonald 1947b) and useful information on the possibility of relating effective source radiation temperature to "noise-temperature" in such cells may emerge from future extension of this work.

* Ignoring, here, the correction term due to Bose–Einstein statistics.

In the secondary-emission amplifier (electron-multiplier) we have an example of enhanced fluctuation due to co-operative electron behaviour. In the final stage a group of n electrons (say $n \simeq 100$–$1{,}000$) will be released effectively simultaneously as a consequence of the transit of a single primary electron in the first stage. The effective quantum of charge in the output current is then ne and a large part of the noise generated is to be attributed to the fact that the simple shot equation for such a current must now be written $\overline{\delta I^2} = 2(ne)I\,df$. An excellent analysis of noise in electron multipliers is due to Shockley and Pierce (1938)*. A general feature of all fluctuation phenomena is that we should be enabled to deduce from experimental data further information about the nature of the physical processes involved or the validity of the theoretical foundations. Secondary emission forms no exception and Ziegler (1936 a, b) and others have found for instance that the simple Poisson law frequently assumed for the distribution of emitted secondaries per primary is inadequate to account for the observations. Hayner and Kurrelmeyer (1935, 1937) have been able to correlate a large amount of data by assuming that the primaries could be divided into three groups: "reflected" electrons, "buried" electrons (producing no secondaries), and those producing true secondaries for which then the assumption of Poisson's law was found to suffice. A more formal statement would appear to be a selection of the cases of $n_s/n_p = 0$, 1 as singular cases.

3.3. *Low-frequency Noise in Valves and Semiconductors etc.*

Over the past twenty years or so sources of noise additional to those (shot and thermal) generally regarded as of a "fundamental" nature have been observed in semiconductors and thermionic valves. Characteristic is the degree of dependence on the current magnitude and on an inverse power of the frequency. An immediate consequence of the latter characteristic is that with increase of frequency these sources ("current noise" in resistances, "flicker noise" in valves) may be subordinated to the true shot and/or thermal fluctuation; measurements of the latter are therefore always (now) conducted at frequencies well removed from the audio range (often in the order of a megacycle per second). On the other hand, there are several fields of investigation—physiological amplifiers (*vide infra*) and ionizing particle-counters for example—where very low frequency stimuli must be recorded. As a result interest in the subject, previously somewhat sporadic, has rather recently revived; a conference on these topics † in 1947 provided a considerable amount of valuable information.

Following a general introductory review by R. A. Smith, E. J. Harris presented a survey of experimental data, mainly around 30 c/s., on semiconducting resistances ‡ (pyrolytic film and carbon composition) and triode and pentode valves (with oxide-coated emitters). In the absence of current only thermal noise (Nyquist's law) was observed, as to be expected, in the resistors; when as little as 1 μa. current flowed, however, the noise intensity rose to several times this value. The general results agree fairly well with a dependence of the "anomalous" noise on the square of the current and on the inverse first power of the frequency; in the case of valves, however, below 300 c/s. the variation between members

* Pierce (1948) has recently published a review in simplified terms of shot and thermal noise including a discussion of secondary-emission amplifiers. The writer, however, must admit to finding the "elementary" treatment rather more difficult to understand than Shockley and Pierce's original analysis.

† Conference on Flicker Noise—papers unpublished, but quoted here with individual authors' permission.

‡ In the range ~100 kΩ–200 kΩ.

of the same type was as great as between different specific types (see also Harris and Bishop 1948). Data on vacuum photocells and secondary-emission photomultipliers was given by D. A. H. Brown; the result of greatest importance certainly seems to be the absence of flicker effect in photocells. Brown suggested that flicker effect might have been expected in photocells because of the semiconducting nature of their emissive surfaces (the cells examined by him had Ag–O–Cs and Sb–Cs cathodes). To the writer, however, this does not seem immediately relevant unless some relatively slow variation of photo-emission probability (cf. Smith 1948) were specifically associated with the semiconductor characteristics. Some general considerations of low-frequency amplifier design based on the available data were discussed by A. B. Gillespie (see also Gillespie, unpublished).

A very valuable contribution by G. G. Macfarlane, since published elsewhere (1947), developed a theory of low-frequency noise intended to be of general application to all types of semiconducting element including the thermionic valve with oxide cathode where the semiconducting surface is then regarded as the cathode–vacuum interface. Schottky (1926) first derived an expression for the excess flicker noise: $\overline{\delta I^2} \propto I^2/(\omega^2+\alpha^2)$, where $\omega = 2\pi f$ and $1/\alpha$ is a time-constant characteristic of the effect—this he visualized as a random variation in density of foreign atoms or molecules on the surface of the cathode producing relatively large, local variations in emissivity. Schottky's theory followed Johnson's discovery (1925) of the phenomenon, reported in an excellent paper which should certainly be studied by anyone interested in the field. Schottky's expression agreed quite well as regards the dependence on current with Johnson's results; the frequency variation, however, was less satisfactory and Moullin (1938), for example, pointed out that Johnson's data indicated more nearly a variation: $\overline{\delta I^2} \propto I^2/\omega^{1\cdot2}$. Macfarlane's theory may be regarded as a considerable development of Schottky's model; metallic atoms from the body of the emitter diffuse under a concentration gradient "up" to the surface in relatively isolated clusters, are there ionized and are thus conducted back by the electric field subsequently into the body of the emitter. While the metallic atoms (e.g. barium in the case of a BaO cathode) are on the surface they are assumed to be strongly polarized and the effective work function is then reduced there by an amount directly proportional to the strength of the dipolar layer. The thermionic emission can thus be related directly to the number of Ba adatoms at one of the localized surface points where "eruption" has taken place. By calculation of the correlation-function for the emission current and thence the power-spectrum, Macfarlane derives a law of the form $\overline{\delta I^2} = I^{x+1}/\omega^x \quad (1<x<2)$.

Rather encouraging agreement is found with a considerable range of experimental data on low-frequency noise. Lead-sulphide cells and carbon resistors (Harris, Abson and Roberts 1947, Christensen and Pearson 1936) are also found to obey a law of this type suggesting that a similar mechanism is responsible in these cases. The primary cause of flicker-noise in tungsten-filament valves (Johnson 1925, Milner—Conference unpublished) appears still uncertain. Further research would seem very desirable to determine definitely whether such valves could be of value in low-noise L.F. amplifiers. One particular difficulty is the separate identification and elimination of microphonics (arising from mechanical resonances in the structure).

In the field of physiology particularly, galvanometer amplifiers seem to have great promise; only thermal Brownian movement is to be expected in a galvanometer as an evitable source of fluctuation. The elimination of extraneous mechanical vibration, however, presents a problem comparable with that of microphony in valves. A. V. Hill (1934, 1938, 1948) has used galvanometer amplifiers of special design with very great success in muscle studies; Frankenhaeuser and the writer (1948) have designed a modified type of amplifier for use in connection with nerve-potential recording: an essential feature is the incorporation of a high degree of negative feedback to provide stability, augmented frequency-response and high input-impedance.* A point of considerable interest is that while a number of nerve-responses (e.g. to brushing and pressing stimuli) have been clearly identified in electro-physiological records, the pain-response has not yet been observed in a single fibre. Figures 6(*a*) and 6(*b*) indicate the importance of noise-reduction in amplifiers for this field.

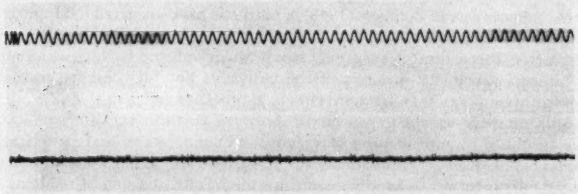


Figure 6(*a*). Weak action-potentials from rabbit's sural nerve produced by normal cutaneous touching stimulus [timing-wave 50 c/s.]
[Courtesy of B. Frankenhaeuser and G. Weddell, Oxford.]

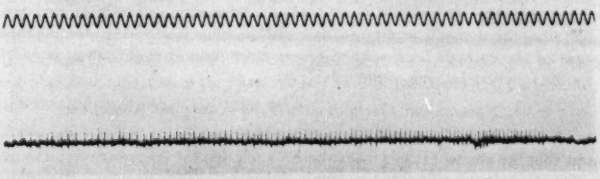


Figure 6(*b*). Action-potentials from rabbit's sural nerve arising from slow-accommodation touch receptor.
[*Vide* B. Frankenhaeuser, *J. Physiol.* (in press).]

The Dutch physicists, Milatz, Bloembergen *et al.* (1946, 1948) have developed over a number of years a photoelectric A.C. galvanometer amplifier. It is claimed that an accuracy equal to that of an ideal D.C. amplifier is achieved; their amplifier, however, is not sensitive to external disturbances because of its essentially resonant character (*vide infra*) or to battery drift, and can be built with ordinary radio components. The galvanometer electromagnet is fed with 50 c/s. current and the instrument then only responds if the current injected into the coil is of

* The output-impedance of a rabbit's nerve may be $\sim 50\ \text{k}\Omega$.

the same frequency and of the appropriate phase. The photocell illumination is interrupted by means of a rotating disc driven from a synchronous motor from the same 50 c/s. source as that for the magnet. A band-width of approximately 1 c/s. occurs naturally if the galvanometer time-constant is approximately 1 sec. The "dark" thermal-emission of the Cs photocell is practically eliminated by cooling in liquid air and high overall sensitivity results. The essential principle has much in common with the Dicke method for measuring fluctuation or radiation of low level. Say we wish to examine the noise received from the sun on a microwave receiver. If a waveguide antenna be exposed to the sun the noise detectable at the receiver output may only differ very slightly from that observed under normal conditions. Let us, however, now "interrupt" this radiation (say by some absorbent shutter) at perhaps 30 c/s.; the small *difference* between the effective solar and terrestrial noise will then essentially be amplifiable as a 30 c/s. "signal" of extremely small band-width.

Closely allied to this field is the problem of the limiting sensitivity of electrometers. Recently van Zolingen (1948) in his thesis has considered the application of differential negative feedback (through a photoelectric amplifier) to a Hoffmann electrometer, the natural air-damping being much reduced by high evacuation. The ultimate sensitivity in terms of the inevitable but high leakage resistance of the instrument can then be more closely approached (cf. also v. d. Ziel 1942). van Zolingen discusses the improvement in overall performance after application of the feedback by introduction of an effective temperature, and in this sense apparent temperatures of approximately 2° K. are achieved. The writer deprecates this mode of performance specification since there is no question of reducing the thermal fluctuation below the fundamental limit set by Nyquist's formula in terms of the inherent instrument resistance. It would appear more logical to specify the improvement in efficiency arising consequent upon the alteration in effective dynamic impedance and frequency response of the system due to the feedback. The thermionic valve when operated as a cathode-follower may similarly improve the effective signal/noise performance of a high-impedance source, since the dynamic input impedance of the value is very considerably raised because of the cathode feedback. Correspondingly, however, a limit would be set by the inherent leakage of the valve.

3.4. *Spontaneous Fluctuations of Temperature and Radiation*

In deriving the expression for the electrical noise of a passive element we assume that the element is in a "heat-bath" such that the (macroscopic) temperature of the body may be regarded as strictly constant. When, however, considering elements which, by virtue of their small heat-capacity (and high thermo-electrical characteristics *) may be used for the detection of, say, infra-red radiation it becomes necessary to consider the spontaneous temperature fluctuations of the body as a whole. A bolometer, for example, operates by virtue of the high resistance-temperature characteristic of certain materials. A steady current is passed through it from an external source and the change of electrical resistance on reception of radiation is thus converted into an electrical potential difference. It might occur in practical cases that the electrical fluctuation arising from the temperature fluctuations would override the Johnson noise of the element when placed at the input of a sensitive amplifier. It is known (cf. Fürth

* Using the term in a general sense.

1920a) that the total thermal energy fluctuation is given by $\overline{\delta E^2} = kT^2C$ or $\overline{\delta T^2} = kT^2/C$, where C is the thermal capacity of the element.

To utilize the results we require, however, to know the distribution in frequency; Milatz and Van der Velden (1943) and Daunt and Kompfner (1945–46) have derived such expressions and thence discussed the optimum design of such thermal detectors. Further contributions have been made by Kompfner and others (e.g. MacDonald 1947).

The voltage fluctuations of a conventional antenna may be treated either through a direct application of Nyquist's theorem in terms of the radiation resistance or else by a consideration of the equilibrium radiation interchange with the surrounding medium (Burgess 1941, 1946b); exactly similarly the energy fluctuations of a thermal receiver may also be analysed from the standpoint of fluctuations in thermal radiation. This analysis has been carried out by Lewis (1947) and Daunt and Kompfner based on the classical work of Einstein (1909) and Planck (1924) leading to results in concordance with the other mode of approach.

Radiation fluctuations may readily be appreciated in terms of random variations in the number of photons arriving, although Bose–Einstein rather than classical statistics are applicable. The lower limit of visual acuity is naturally of considerable interest in this connection; Pirenne (1947) has recently reviewed the problem and it appears that many fundamental physiological properties of the eye are explicable through the quantization of radiation and fluctuations in number of quanta received by individual receptors. de Vries (1943) (cf. also Pirenne 1946) has shown that the differential sensitivity of the eye is nearly equal to the statistical fluctuations in the number of quanta absorbed in the retina. Pirenne (1947) considers the relative frequency of perception of a very faint image in relation to the (nominal) stimulus and concludes that the observed variability of limiting visual response is explicable by random fluctuations in the photon arrival-rate without necessarily calling for any additional physiological hypothesis.

de Vries (1948) has recently reviewed and discussed the aural threshold, in relation to the Brownian movement of the surrounding air, of the vibrating system of the ear and of the sense-cells. He concludes that a discrepancy exists on present theories of hearing between the observed low audible threshold and the magnitude of the thermal agitation of the sense cells.

REFERENCES

Bakker, C. J., 1941, *Physica*, **8**, 23.

Barnes, R. B., and Silverman, S., 1934, *Rev. Mod. Phys.*, **6**, 162.

Bell, J., 1946, *J. Instn. Elect. Engrs.*, **93**, (III A), 187.

Bennett, W. R., and Rice, S. O., 1934, *Phil Mag.*, **18**, 422.

Bleaney, B., Ryde, J. W., and Kinman, T. H., 1946a, *J. Instn. Elect. Engrs.*, **93**, (III A), 184; 1946b, *Ibid*, **93**, (III A), 847; 1947, *Ibid., Discussion*, **94**, (III), 339.

Bloembergen, N., Purcell, E. M., and Pound, R. V., 1948, *Phys. Rev.*, **73**, 679.

Born, M., 1926, *Z. Phys.*, **37**, 863.

Bouligand, G., 1934, *Relations d'Incertitude en Géometrie et en Physique—Act. scient. et indus.* (Paris).

Bradfield, G., Bartlett, J. G., and Watson, D. S., 1946, *J. Instn. Elect. Engrs.*, **93**, (III A), 128.

Brown, R. C., 1938, *Rep. Prog. Phys.*, Vol. V (London : Physical Society).

Burgess, R. E., unpubl. series of reports through D.S.I.R. on analysis of noise through non-linear devices etc.; 1941, *Proc. Phys. Soc.*, **53**, 293; 1943, *Wireless Engr.*, **20**, 66; 1945, *Ibid.*, **22**, 56; 1946a, *Ibid.*, **23**, 217; 1946b, *Proc. Phys. Soc.*, **58**, 313.

CAMPBELL, N. R., 1909, *Proc. Camb. Phil. Soc.*, **15,** 117, 310; 1910, *Ibid.*, **15,** 513; 1910, *Z. Phys.*, **11,** 826.

CAMPBELL, N. R., and FRANCIS, V. J., 1946, *J. Instn. Elect. Engrs.*, **93,** 45. (See also unpublished work from G.E.C.)

CAMPBELL, N. R., FRANCIS, V. J., and JAMES, E. G., 1945, *Wireless Engr.*, **22,** 333; 1946, *Ibid.*, **23,** 16, 74, 116.

CHANDRASEKHAR, S., 1943, *Rev. Mod. Phys.*, **15,** 68.

CHRISTENSEN, C. J., and PEARSON, G. L., 1936, *Bell Syst. Tech. J.*, **15,** 197.

DAUNT, J. G., and KOMPFNER, R., 1945–46, Series of unpublished reports from Clarendon Laboratory on thermal detectors etc.

DOMB, C., 1946, *Proc. Camb. Phil. Soc.*, **42,** 245.

EHRENFEST, P., 1927, *Z. Phys.*, **45,** 455.

EINSTEIN, A., 1905, *Ann. Phys. Lpz.*, **17,** 549; 1907, *Z. Elektrochem.*, **13,** 41; 1908, *Ibid.*, **14,** 235; 1906, *Ann. Phys. Lpz.*, **19,** 289, 371; 1909, *Phys. Z.*, **10,** 185, 817; 1911, *Ann. Phys. Lpz.*, **34,** 591; 1926, *Investigations on the Theory of the Brownian Motion* (collected papers, edited by R. Fürth) (London : Methuen).

ELLICKSON, R., 1938, *Phys. Rev.*, **54,** 572.

FEATHER, N., 1943, *Proc. Camb. Phil. Soc.*, **39,** 84.

FOSTER, J., 1946, *J. Instn. Elect. Engrs.*, **93,** (III A), 192, 868.

FRANKENHAEUSER, B., and MACDONALD, D. K. C., 1948, *J. Sci. Instrum.* (in press).

FRÄNZ, K., 1940, *Elekt. Nachr.-Tech.*, **17,** 215 (addend. and corrig., 1942, *Ibid.*, **19**); 1941, *Hochfrequenztech. u. Elektroakust.*, **57,** 146; 1942, *Elek. Nachr.-Tech.*, **19,** 166.

FRIIS, H. T., 1944, *Proc. Inst. Radio Engrs.*, **32,** 419.

FRY, T. C., 1925, *J. Franklin Inst.*, **199,** 203.

FURRY, W. H., 1937, *Phys. Rev.*, **52,** 569.

FÜRTH, R., 1920a, *Schwankungserscheinungen in der Physik* (Brunswick : Vieweg); 1920b, *Z. Phys.*, **2,** 244; 1948a, *The natural limit of accuracy of observation—Science Progress*, No. 143 (see also 1933, *Z. Phys.*, **81,** 143; 1948b, *Proc. Roy. Soc.* A, **192,** 593.

FÜRTH, R., and MACDONALD, D. K. C., 1946, *Nature, Lond.*, **157,** 807; 1947a, *Proc. Phys. Soc.*, **59,** 388; 1947b, *Nature, Lond.*, **159,** 608.

GILLESPIE, A. B., *The Design of A.C. Amplifiers for use with ionization chambers*: *An introduction to A.C. Amplifiers for Nuclear Physics Research* (unpublished).

DE HAAS-LORENTZ, G. L., 1913, *Die Brownsche Bewegung und einige verwandte Erscheinungen* (Brunswick : Vieweg).

HARRIS, E. J., ABSON, W., and ROBERTS, W. L., 1947 (unpublished).

HARRIS, E. J., and BISHOP, P. O., 1948, *Nature, Lond.*, **161,** 971.

HAYNER, L. J., and KURRELMEYER, B., 1935, *Physics*, **6,** 323; 1937, *Phys. Rev.*, **52,** 952.

HILL, A. V., 1934, *J. Sci. Instrum.*, **11,** 309; 1938, *Proc. Roy. Soc.* B, **126,** 136; 1948, *J. Sci. Instrum.*, **25,** 225.

HOPKINSON, R. G., 1946, *J. Instn. Elect. Engrs.*, **93,** (III A), 795.

JOHNSON, J. B., 1925, *Phys. Rev.*, **26,** 71.

KHINTCHINE, A., 1934, *Math. Ann.*, **109,** 604.

KLEEN, W., 1941, *Die Telefunkenröhre*, **23,** 274.

KLEMENS, P. G., 1948, *Phil. Mag.*, **39,** 656.

KOMPFNER, R., 1947a, *Proc. Inst. Radio Engrs.*, **35,** 124; 1947b, *J. Brit. Instn. Radio Engrs.*, **7,** 117; 1947c, *Wirelss Engr.*, **24,**; Unpublished reports on radiation fluctuations and temperature fluctuations from Clarendon Laboratory.

KOMPFNER, R., HATTON, J., SCHNEIDER, E. E., and DRESEL, L. A. C., 1946, *J. Instn. Elect. Engrs.*, **93,** (III A), 225.

LANGEVIN, M. P., 1908, *C.R. Acad. Sci., Paris*, **146,** 530.

LEWIS, W. B., 1947, *Proc. Phys. Soc.*, **59,** 34.

MACDONALD, D. K. C., 1944, *Phil. Mag.*, **35,** 386; 1947, Unpublished report from Clarendon Laboratory on temperature fluctuations analysis; 1948a, *Proc. Roy. Soc.* A, **195,** 225; 1948b, *Research*, **1,** 194; 1949a, *Proc. Camb. Phil. Soc.* (in press); 1949b, *Phil. Mag.*, **40,** 561.

MACDONALD, D. K. C., and KOMPFNER, R., 1949, *Proc. Inst. Radio Engrs.* (submitted for publication).

MACFARLANE, G. G., 1947, *Proc. Phys. Soc.*, **59,** 366.

MILATZ, J. M. W., and BLOEMBERGEN, N., 1946, *Physica*, **11,** 449.

MILATZ, J. M. W., ENDT, P. M., ALKEMADE, C. TH. J., and OLINK, J. TH., 1948, *Physica*, **14**, 260.
MILATZ, J. M. W., and VAN DER VELDEN, H. A., 1943, *Physica*, **10**, 369.
MOFFATT, J., 1946, *J. Instn. Elect. Engrs.*, **93** (III A), 228.
MOULLIN, E. B., 1938, *Spontaneous Fluctuations of Voltage* (Oxford: Clarendon Press), chap. VI, p. 165.
NIESSEN, K. F., and BAKKER, C. J., 1938, *Physica*, **5**, 977.
NORTH, D. O., 1942, *R.C.A. Rev.*, **6**, 332.
NORTH, D. O., and FERRIS, W. R., 1941, *Proc. Inst. Radio Engrs.*, **29**, 49.
NORTH, D. O., *et al.*, 1940–41, *R.C.A. Rev.*, **4**, 269, 441; **5**, 106, 224, 371, 505; **6**, 114.
NORTON, K. A., and OMBERG, A. C., 1947, *Proc. Inst. Radio Engrs.*, **35**, 4.
NYQUIST, H., 1927, *Phys. Rev.*, **29**, 614; 1928, *Ibid.*, **32**, 110.
ORNSTEIN, L. S., 1919, *Proc. Roy. Acad. Amst.*, **21**, 96.
ORNSTEIN, L. S., and MILATZ, J. M. W., 1938, *Physica*, **5**, 971.
PAWSEY, J. L., 1935, *Proc. Camb. Phil. Soc.*, **31**, 125.
PEARSON, K., 1905a, *Nature, Lond.*, **72**, 294; 1905b, *Ibid.*, **72**, 342.
PIERCE, J. R., 1947, *Proc. Inst. Radio Engrs.*, **35**; 1948, *Bell Syst. Tech. J.*, **27**, 158.
PIRENNE, M. H., 1946, *Proc. Camb. Phil. Soc.*, **42**, 78; 1947, *Science Progress*, No. 140, 605.
PLANCK, M., 1924, *Ann. Phys. Lpz.*, **73**, 272.
RAGAZZINI, J. R., 1942, *Proc. Inst. Radio Engrs.*, **30**, 277.
RATCLIFFE, J. A., 1948, *Nature, Lond.*, **162**, 9.
RATCLIFFE, J. A., and PAWSEY, J. L., 1933, *Proc. Camb. Phil. Soc.*, **29**, 301.
RAYLEIGH, LORD, 1905, *Nature, Lond.*, **72**, 318.
RICE, S. O., 1944, *Bell Syst. Tech. J.*, **23**, 282; 1945, *Ibid.*, **24**, 46; 1948, *Ibid.*, **27**, 109.
RIVLIN, R. S., 1945, *Phil. Mag.*, **36**, 688.
ROWLAND, E. N., 1936, *Proc. Camb. Phil. Soc.*, **32**, 580; 1937, *Ibid.*, **33**, 344.
SCHOTTKY, W., 1918, *Ann. Phys. Lpz.*, **57**, 541; 1926, *Phys. Rev.*, **28**, 74.
SCHOTTKY, W., and SPENKE, S., 1937, *Wiss. Veröff. Siemenswerken*, **16**, 1.
V. SCHWEIDLER, E., 1905, *Über Schwankungen der radioaktiven Umwandlung, Congr. int. Radiol. et Ionis.* (Liége); 1910, *Phys. Z.*, **11**, 614; 1913, *Ibid.*, **14**, 198.
SHOCKLEY, W., and PIERCE, J.R., 1938, *Proc. Inst. Radio Engrs.*, **26**, 321.
SLACK, M., 1946, *J. Instn. Elect. Engrs.*, **93**, 76.
SMITH, R. A., 1948, Discussion on noise papers, Conference on Scientific Radio, 1948, *J. Instn. Elect. Engrs.*
V. SMOLUCHOWSKI, M., 1906, *Ann. Phys. Lpz.*, **21**, 756.
SMYTH, C. N., 1946. *Nature, Lond.*, **157**, 841.
STRONG, J., 1938, *Procedures in Experimental Physics* (New York: Prentice-Hall; London: Blackie), p. 298.
STRUTT, M. J. O., and V. D. ZIEL, A., 1941, *Physica*, **8**, 424; 1942a, *Ibid.*, **9**, 513; 1942b, *Ibid.*, **9**, 528; 1946, *Wireless Engr.*, **23**, 241; 1948, *Proc. Inst. Radio Engrs.*, **36**, 19.
UHLENBECK, G. E., and ORNSTEIN, L. S., 1930, *Phys. Rev.*, **36**, 823.
ULLRICH, E. H., and ROGERS, D. C., 1946, *J. Instn. Elect. Engrs.* (III A), **93**, 233.
VOLZ, H., 1935, *Z. Phys.*, **93**, 539.
DE VRIES, HL., 1943, *Physica*, **10**, 553; 1948, *Ibid.*, **14**, 48.
WANG, M. C., and UHLENBECK, G. E., 1945, *Rev. Mod. Phys.*, **17**, 323.
WARD, A. G., 1942, *Proc. Roy. Soc.* A, **181**, 183.
WIENER, N., 1930, *Act. Math., Stockholm*, **55**, 117.
WILSON, R., 1947, *Proc. Soc. Psych. Res.*, **48**, 213.
WOODWARD, P. M., 1948a, "The Mathematical Description of Random Noise", *T.R.E. Journal*; 1948b, *Proc. Camb. Phil. Soc.*, **44**, 404.
ZIEGLER, M., 1936a, *Physica*, **3**, 1; 1936b, *Ibid.*, **3**, 307.
V. D. ZIEL, A., 1942, *Physica*, **9**, 177; 1947a, *Philips Res. Rep.*, **2**, 321; 1947b, *Nature, Lond.*, **159**, 675.
V. D. ZIEL, A., and VERSNEL, A., 1947, *Nature, Lond.*, **159**, 640; 1948, *Philips Res. Rep.*, **3**, 13.
V. ZOLINGEN, J. J., 1948, *De Brownse Beweging van Electrometers en een methode om deze te vermindern—Diss., Leiden.* (See also papers by Milatz, J. M. W., v. Zolingen, J. J., Wapstra, A. H., and van Iperen, B. B., *Physica*, **14**, 1948, not in print at time of writing.)

RECENT NUCLEAR EXPERIMENTS WITH HIGH VOLTAGE X-RAYS

By W. BOSLEY* and J. D. CRAGGS†

CONTENTS

	PAGE
§ 1. Introduction	82
§ 2. Experimental techniques	83
2.1. The use of Geiger counters and ionization chambers in measurement of 5–20 MeV. X-rays	83
2.2. The use of Wilson chambers with electron accelerators	84
§ 3. Experimental results	
3.1. Photo-fission	86
3.2. Nuclear photo-effect	88
3.3. The (γ, p) reactions	92
3.4. Energy loss of fast electrons and positrons and pair production	96
3.5. Miscellaneous experiments	99
References	100

§ 1. INTRODUCTION

THE purpose of the present Report is to give an account of some of the experiments that have been performed, since about 1941, with the new high energy electron accelerators (betatrons and synchrotrons). For chronological reasons most of the work so far carried out has been with betatrons, but it seems likely that as electron energies higher than the 100 MeV. now available (Westendorp and Charlton 1945) are reached, a much higher proportion of the researches will be made with synchrotrons, several of which (at the University of California, at the Massachusetts Institute of Technology, and, in Britain, for the University of Glasgow) are now being built for 300 MeV. Linear accelerators for electrons and for protons have been, or are being, built in various laboratories; the former appear to be capable of giving very high X-ray outputs, and fuller accounts of their performance are awaited with much interest (see Fry and Walkinshaw 1949).

It is not proposed to describe the accelerators themselves in this Report, but the following review and original papers may be consulted for detailed references: betatrons are described generally by Kerst (1946), Bosley (1946), and others, and synchrotrons by Elder *et al.* (1947) and Fry *et al.* (1947).

It has been necessary for completeness to refer in several places to experiments with high energy γ-rays, which have been used for photo-disintegration work (Bothe and Gentner 1939 and others), for experiments on electron scattering, pair production etc., and for the (γ, p) reaction. We do not propose to define the term "high-voltage" used in the title of this Report, since mention has been made of certain experiments with Van de Graaff generators (see § 3) on, for example, the photo-disintegration of Be and D, but the lower energy limit considered generally in this paper is approximately 5 MeV.

It is clearly impossible in a short review to cover adequately all the fields of application for high voltage X-rays. The Report deals only with nuclear physical experiments carried out with high energy electron accelerators; the work on radiography and on the preliminary investigations of the biological and clinical effects of high energy radiations have not been discussed.

* Department of Natural Philosophy, University of Glasgow.
† Department of Electrical Engineering, University of Liverpool.

§ 2. EXPERIMENTAL TECHNIQUES

2.1. *The Use of Geiger Counters and Ionization Chambers in Measurement of* 5–20 *Mev. X-rays*

The purpose of this and the following section is to indicate briefly some special aspects of certain measurement techniques. In general, standard experimental methods applicable to work with, and measurements of, betatron and synchrotron X-ray beams often need modification for two major reasons, viz. the high quantum energies and the peculiar pulsed character of the beams, which generally consist of a regular series of short pulses, lasting for perhaps one or a few microseconds, separated by relatively long intervals ($\sim$1/100–1/10 sec.)

Ionization chambers have been used by several authors (e.g. Kerst 1942, 1946) to give the mean value in time of the radiation flux, and Elder *et al.* (1947) have used photomultiplier tubes, which should enable the beam pulse-shape to be followed accurately. Geiger counters have, of course, been used to measure the activities of specimens irradiated and rendered radioactive in the beam.

In cases where Geiger (non-proportional) counters are used in the beam it is necessary to allow for the finite dead-time ($\sim$100 μsec.) of such devices. Muehlhause and Friedman (1946) drew attention to this limitation and pointed out that as the beam intensity is increased from zero the counting rate varies from zero up to the pulse recurrence rate M for the accelerator ($\sim$180 sec.). The counting rate is given by $N = M\{1-\exp(-N_0\overline{E})\}^{\alpha}$ where N_0 is the number of quanta, per beam pulse, falling on the counters, each of efficiency $\overline{E}$, and α is the number of counters used in coincidence. It is advantageous to reduce $\overline{E}$ and α. Work on $\overline{E}$ has been carried out by, for example, Yukawa and Sakata (1937) and Dunworth (1940) to study high-energy γ-radiations and to investigate the relative importance of the photoelectric effect, Compton recoils, and pair production in determining $\overline{E}$.

The above expression implies that the counter dead-time is much longer than the pulse duration. So far as the present authors know, the use of proportional or sub-threshold Geiger counters has not been explored with short pulses of radiation. Muehlhause and Friedman (1946) have counted at rates of 100,000/sec. using counters in which the full Geiger pulse is not allowed to develop (cf. Stever 1942).

Photomultipliers do not suffer from the particular limitation due to dead-time: Elder *et al.* (1947) have published a photograph of the wave-form of the X-ray pulse from the Schenectady 70 MeV. synchrotron, taken with an R.C.A. 931 photomultiplier tube. The technique is clearly capable of extension to many other related problems.

The use of ionization chambers for the measurement of X-ray beam strengths in the region 5–20 MeV. entails the extension of existing techniques. The theory of air-wall chambers for measurements of X-ray beam strengths in roentgens has been discussed in many places (Gray 1928 and Laurence 1937), and an authoritative review is given in a book by Wilson (1945). No description of the use of air-wall chambers for 20 MeV. beams has been published, although experiments have, no doubt, been carried out in several laboratories; for example, the present authors have constructed a series of graphite-walled chambers (5 cm^3 volume) of wall thickness varying from about 0·4 to 4 cm.

Recent and comprehensive reviews of modern electronic techniques for ionization chambers have been given by Lewis (1942) and by Corson and Wilson (1948). The latter authors briefly discuss some of the latest methods using electron collection which might prove of great value in short-pulse work with electron accelerators.

Lax (1947) has considered in great detail the behaviour of a graphite chamber having two walls each 4·5 cm. thick separated by a thin gas layer. 25 MeV. electrons would not penetrate such walls, but above this energy the chamber does not behave as a normal ionization chamber, and its sensitivity will be somewhat reduced. 4·5 cm. of graphite is equivalent to about one-fifth of a radiation length (which is roughly the distance in which the energy of a fast particle is reduced to $1/e$ of its original value by radiative processes (see also Rossi and Greisen 1941), and the relative sensitivity of the chamber, S/G, as a function of x-ray (or γ-ray) energy is given in Figure 1. The spectrum was assumed flat from 0 to U MeV.

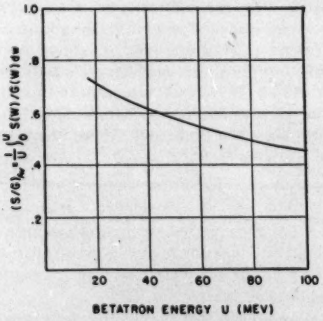


Figure 1. The response of a graphite ionization chamber to high energy x-rays.

This arbitrary flat spectrum corresponds approximately to a 100 MeV. bremsstrahlung spectrum (Heitler 1944, p. 170) which would presumably be obtained with a thin target (see below, §3.4). For a 20 MeV. beam this approximation would probably not be justified as the thin target spectrum tends to a triangular shape (Figure 2). Lax enumerates in great detail the minor corrections due to attenuation in the chamber walls, electron scattering in the latter, the electronic radiation losses and non-linearity of the range–energy relation. For 20–100 MeV. the combined effect of these corrections is not more than $\pm 7\%$.

2.2. *The Use of Wilson Chambers with Electron Accelerators*

Wilson chambers have been used with betatrons in several experiments. In certain cases it is convenient, for example, to avoid unnecessary and wasteful operation of the betatron magnet etc. during the long ($\sim$1 min.) waiting periods inevitable with Wilson chambers, to excite the magnet in short bursts, synchronized with the Wilson chamber cycle. Thus, Baldwin *et al.* (1947) describe electronic methods for controlling the injection of electrons in the betatron, whether the magnet be pulsed or excited continuously. The object of this work was essentially

to ensure the production, in the Wilson chamber, of sharp uniform tracks formed at corresponding instants in every expansion chamber cycle.

The sequence of events is as follows: The primary signal from the Wilson chamber first produces pulse excitation of the main magnet (unless continuously excited), then causes the injection of one burst of electrons. The flashing of the Wilson chamber illuminating system can also conveniently be controlled by the same master circuit and, finally, the delays between (*a*) expansion and (*b*) the passage of the beam through the chamber, and between (*b*) and the illumination of the tracks for photography are controlled.

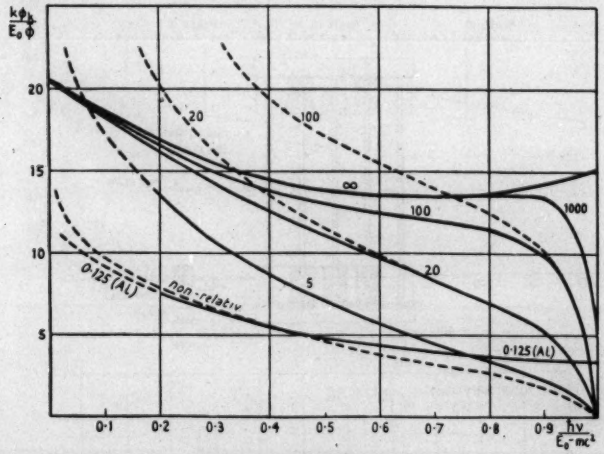


Figure 2. Bremsstrahlung spectra.

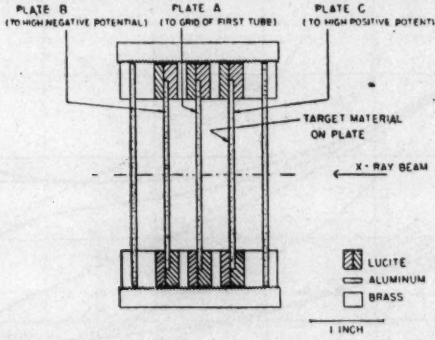
Koch and Robinson (1948) pointed out that pulsing of the main magnet windings with a magnetic contactor (Baldwin *et al.* 1947) is sometimes undesirable since the magnet current depends to some extent on the rate of closing of the contacts, and on their condition. Koch and Robinson excited the winding of their 22 MeV. betatron magnet by discharging the condenser bank, via two ignitrons (type GL. 415) in series, through the main field coil. One of the ignitrons was fired by a pulse produced by the cloud-chamber timing system, and the electron injection process could be controlled by any of the usual methods which, since the instant of injection must be accurately timed with reference to the magnet field, were necessarily triggered by a pulse developed in a peaker-strip winding (Kerst 1942) or pick-up coil in the main magnetic circuit.

The authors have used a different electronic system for use with a Wilson chamber. For some experiments now in progress it is necessary to fire the X-ray beam through the chamber inside its sensitive time. The electron injector in the betatron (20 MeV.) is arranged to give not only one, but a controllable number of pulses (up to 10) when fired by a control circuit tripped from the Wilson chamber (Heymann and Bosley 1949).

§3. EXPERIMENTAL RESULTS

3.1. *Photo-Fission*

Very soon after the discovery of the fissile properties of uranium, Roberts *et al.* (1939) and Heyn *et al.* (1939) reported unsuccessful attempts to produce fission in uranium and thorium by bombardment with the γ-rays from lithium and from fluorine. Haxby *et al.* (1941), using fluorine (6 MEV.) γ-rays, later succeeded in detecting uranium and thorium fission, the cross sections being estimated to be $\sigma_U = 3{\cdot}5 \pm 1{\cdot}0 \times 10^{-27}$ cm^2 and $\sigma_{Th} = 1{\cdot}7 \pm 0{\cdot}5 \times 10^{-27}$ cm^2. An ionization chamber and amplifier were used to detect the fission process in this



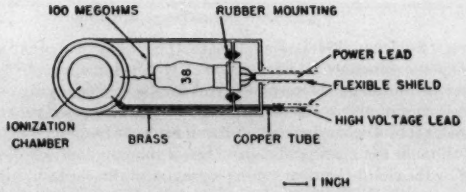


Figure 3. Ionization chamber used in work on photo-fission.

experiment and the disintegrations were counted visually on the screen of a cathode-ray oscillograph, several tests being made to ensure that the phenomena observed were, in fact, due to γ-rays and not to neutrons.

Photo-fission of uranium and thorium has also been investigated by Baldwin and Klaiber (1947), using betatron X-rays of from 10 to 100 MEV. peak energy. In this experiment a differential type of ionization chamber, as shown in Figure 3, was used. This, when properly balanced, was practically insensitive to X-rays. Uranium—or thorium—oxide was coated upon one electrode of the chamber, the signals from which, after passing through a linear amplifier, were applied to a

discriminator and a recording circuit. A monitor oscillograph was also used to check the operation of the apparatus. The incident x-radiation from the betatron was monitored by a Victoreen thimble ionization chamber encased in

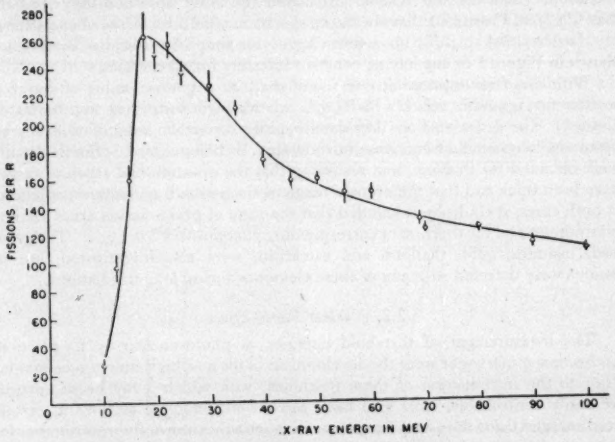


Figure 4. Variation of photo-fission activity of uranium with x-ray energy.

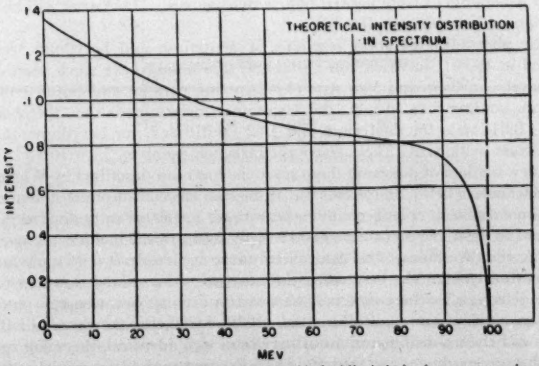


Figure 5. Arbitrary betatron radiation spectrum (dashed line) derived as an approximation to a bremsstrahlung spectrum (full line).

$\frac{1}{8}$ in. lead and placed behind the fission chamber. Measurements were made of the number of fissions produced per roentgen, as measured by the Victoreen chamber, and this was plotted against the peak x-ray energy, a curve of the shape shown in Figure 4 being obtained in the case of uranium.

In order to determine from these results the variation of cross section of the photo-fission reaction with x-ray energy, it is necessary to know (*a*) the spectrum of the betatron radiation and (*b*) the response of the Victoreen chamber to this radiation. Baldwin and Klaiber calculated the latter approximately (see also Lax 1947) and assumed that the x-ray spectrum would be of the shape derived by Heitler (1944, p. 170) for a thin target; for simplicity this was modified as shown in Figure 5 to one having constant intensity for all energies.

With these assumptions it was found that the maximum value of the cross section for uranium was 6×10^{-26} cm^2, this value occurring at approximately 16 MeV. The cross section decreased rapidly on either side of this energy, becoming very small at energies above 30 MeV. or below 8 MeV. Similar results were obtained for thorium, and assuming that the uranium and thorium targets were both thick and that the average range of the fission fragments was the same in both cases, it could be concluded that the ratio of photo-fission cross sections for uranium and for thorium at corresponding energies was $2{\cdot}0 \pm 0{\cdot}1$. Tungsten, lead, bismuth, gold, thallium and samarium were also investigated, but no fissions were detected with any of these elements during long irradiations.

3.2. *Nuclear Photo-Effect*

The measurement of threshold energies in photo-nuclear (γ, n) reactions has become much easier since the development of the new high energy accelerators. Prior to the introduction of these machines, with which x-ray beam energies are easily controllable, work had been carried out on cross sections at certain fixed energies using the γ-rays from elements such as lithium or fluorine bombarded by protons, and also on the disintegration of light elements (deuterium and beryllium) whose thresholds were within the energy range available from Van de Graaff generators or from natural radioactive sources. This work will be briefly summarized here for completeness.

The photo-disintegration reactions of deuterium and beryllium were discovered in 1934 (Chadwick and Goldhaber 1934), and have since been studied extensively. Myers and Van Atta (1942), using paraffin and boron trifluoride neutron counters to detect the reactions, determined the thresholds as $2{\cdot}183 \pm 0{\cdot}012$ MeV. for deuterium and $1{\cdot}62 \pm 0{\cdot}010$ MeV. for beryllium, the ratio of the cross sections at 15 kev. above the threshold being $\sigma_{Be}/\sigma_D = 10/1$.

A new method of detecting these reactions has been described by Wiedenbeck and Marhoefer (1945), and makes use of the resonance absorption of neutrons by rhodium or silver. Using counters with these metals as cathodes the authors obtained straight line plots of neutron activity against peak incident x-ray energy, enabling the thresholds to be determined more accurately by extrapolation than had previously been the case when the neutrons were slowed down in paraffin before detection by the counters; the neutron activity–beam energy plots then gave a power law curve. Collins *et al.* (1939) found that the thresholds for the photo- and electro-disintegrations of beryllium were identical, the value agreeing with that given above, and activity was increased tenfold for a given electron current by first converting it into x-rays before bombarding the beryllium. The neutron yields from photo- and electro-disintegration of beryllium have also been investigated by Wiedenbeck (1946).

Gibson *et al.* (1947) have determined the angular distribution of the protons from the deuterium reaction using the photographic plate technique which has

also been used in preliminary measurements of the X-ray spectrum of a betatron (see §3.4), and Wilson *et al.* (1948) have measured the cross section at 2·62 MeV. to an accuracy of ±6% using an ionization chamber method. The value they found was $16{\cdot}2 \pm 1{\cdot}0 \times 10^{-28}$ cm^2.

Bothe and Gentner in 1939 measured the cross section for the ^{63}Cu (γ, n) ^{62}Cu reaction with lithium and boron γ-rays and compared the yields induced by both radiations in other elements.

The first paper dealing with the nuclear photo-disintegrations produced by a betatron was published by Baldwin and Koch (1945). The (γ, n) reaction was produced in ten elements whose thresholds lay between 9 and 19 MeV., and was detected by the induced radioactivity in the elements (as for example in the reaction $^{14}\mathrm{N} + h\nu \rightarrow {}^{13}\mathrm{N} + \mathrm{n}^1$, $^{13}\mathrm{N} \rightarrow {}^{13}\mathrm{C} + \mathrm{e}^- + \mathrm{e}^+ + Q$). ($Q$ is the binding energy, determined by the threshold value of $h\nu$.)

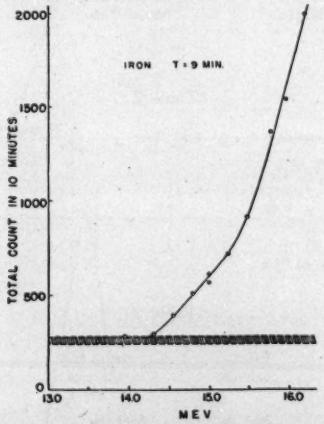


Figure 6. Determination of (γ, n) threshold for iron.

The betatron was fitted with an energy integrating circuit to ensure constant peak output for a given setting of the controls, and irradiation of the element concerned was carried out in a sequence of increasing peak energy values until activity was detected with a β-counter after the irradiation had ceased. A typical curve of activity of the specimen (in this case iron after a 10 min. irradiation) against peak X-ray energy is shown in Figure 6 and the results are tabulated in Table 1.

As the X-ray spectrum of the betatron was not known, cross-section measurements could not be made in this work, and no attempt was made to interpret the shape of the excitation curves, all of which were of the same general shape.

This work was continued by Becker *et al.* (1947) using a 20 MeV. betatron, and the results given in Table 2 were obtained.

Perlman and Friedlander (1948), using the 100 MeV. machine, adopted either the ^{11}C or ^{18}F (γ, n) reactions as standards with which the yields of fifteen other (γ, n) reactions were compared at 50 and 100 MeV., with the results shown in

Table 3, in which the 50 MeV. ^{14}N yield is taken as unity. (Other earlier results for the (γ, n) and (γ, 2n) reactions are given by Baldwin and Klaiber (1946) and are shown in Table 1.)

Table 1

Z	Target element	Active isotope	Half-life (min.)	Threshold (MeV.)
6	Carbon	^{11}C	21	18·7–19·4
7	Nitrogen	^{13}N	10	11·1±0·5
8	Oxygen	^{15}O	2·1	16·3±0·4
26	Iron	^{53}Fe	9	14·2±0·4
29	Copper	^{62}Cu	10·3	10·9±0·3
30	Zinc	^{65}Zn	39	11·6±0·4
34	Selenium	^{79}Se or ^{81}Se	17	9·8±0·5
34	Selenium	^{79}Se or ^{81}Se	57	higher
42	Molybdenum	^{91}Mo or ^{93}Mo	17	13·5±0·4
47	Silver	^{108}Ag	2·3	9·3±0·5
47	Silver	^{106}Ag	24·5	above 9·5

Table 2

Reaction	Period (sec.)	Threshold obs.	Threshold calc.
^{7}Li (γ, p) ^{6}He	1	9·5±0·3	10·1
^{24}Mg (γ, n) ^{23}Mg	11	16·4±0·3	15·5
^{27}Al (γ, n) ^{26}Al	7	14·4±0·3	11·1
^{28}Si (γ, n) ^{27}Si	5	16·9±0·3	16·0
^{32}S (γ, n) ^{31}S	3	15·0±0·3	16·7
^{40}Ca (γ, n) ^{39}Ca	1	16·0±0·3	13·7

Table 3

Parent isotope	Product isotope	Product half-life	Product betas and energies (MeV.)	Relative yield 100 MeV.	Relative yield 50 MeV.
$^{12}_{6}C$	^{11}C	20·5 m.	β+1·04	2·3	2·3
$^{14}_{7}N$	^{13}N	9·9 m.	β+1·2	1·00	1·00
$^{16}_{8}O$	^{15}O	2·1 m.	β+1·7	2·2	2·4
$^{19}_{9}F$	^{18}F	1·87 h.	β+0·7	2·7	2·8
$^{27}_{13}Al$	^{26}Al	7·0 s.	β+2·99	2·3	3·1
$^{31}_{15}P$	^{30}P	2·5 m.	β+3·0	7·2	7·1
$^{35}_{17}Cl$	^{34}Cl	33·0 m.	β+5·1 (80%) β+2·4 (20%)	2·4	2·4
$^{39}_{19}K$	^{38}K	7·6 m.	β+2·5	2·6	2·6
$^{58}_{28}Ni$	^{57}Ni	36 h.	β+0·67	6·3	6·0
$^{63}_{29}Cu$	^{62}Cu	9·9 m.	β+2·6	33	35
$^{69}_{31}Ga$	^{68}Ga	68·0 m.	β+1·9	42	44
$^{71}_{31}Ga$	^{70}Ga	20·0 m.	β−1·7	43	44
$^{110}_{46}Pd$	^{109}Pd	12·7 h.	β−1·1	33	39
$^{109}_{47}Ag$	^{108}Ag	2·33 m.	β−2·8	41	46
$^{121}_{51}Sb$	^{120}Sb	16·6 m.	β+1·53	42	46
$^{187}_{75}Re$	^{186}Re	92·0 h.	β−1·07	85	86

It may be seen from Table 3 that within the accuracy of measurement there is no increase in yield relative to ^{11}C or ^{18}F between 50 MeV. and 100 MeV. (peak), from which one may conclude that either all the cross sections vary in the same way with energy as those for ^{11}C or ^{18}F or that the cross sections above 50 MeV. are very small.

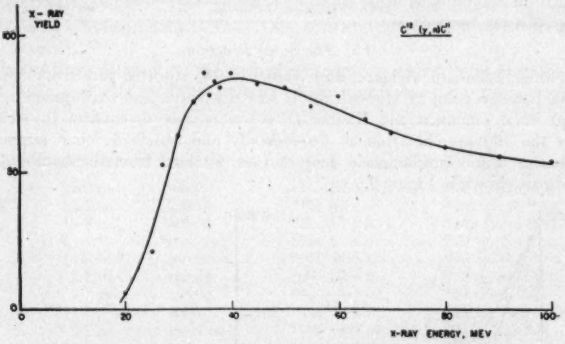


Figure 7. Variation of ^{12}C (γ, n) ^{11}C activity with betatron beam energy.

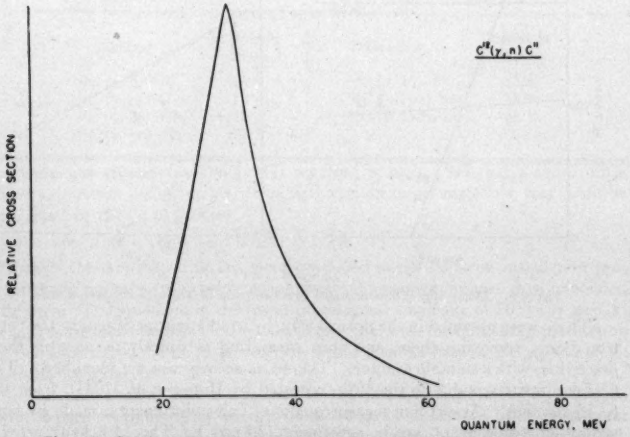


Figure 8. Relative cross-section/quantum energy relation for ^{12}C (γ, n) ^{11}C derived from an assumed simple betatron spectrum.

This latter conclusion was supported by measurements of yields published by Baldwin and Klaiber (1948) using the 100 MeV. betatron. In this work the ^{12}C (γ, n) ^{11}C and ^{63}Cu (γ, n) ^{62}Cu reactions were studied, and with the same

assumptions for x-ray spectrum (and the same calculation for the response of the monitoring Victoreen chamber) as were made in the earlier work on photo-fission (§ 3.1) the yield curves obtained (cf. Figure 7) could be analysed to give the variation of cross section of the reactions with x-ray energy. The result for the carbon reaction is shown in Figure 8. From this curve it may be seen that, if the assumptions made are correct, the reaction is sensitive to only a small range of x-ray energies.

3.3. *The* (γ, p) *Reactions*

Huber, Lienhard, Scherrer and Wäffler (1944) reported an activity obtained in the bombardment of Mg with the 17 MEV. Li γ-rays, and attributable to ^{26}Mg (γ, p) ^{25}Na. Baldwin and Klaiber (1946) were able to confirm the reaction with the 100 MEV. betatron at Schenectady and obtained much supporting evidence. For example, their decay curves obtained from the bombardment of Mg are shown in Figure 9.

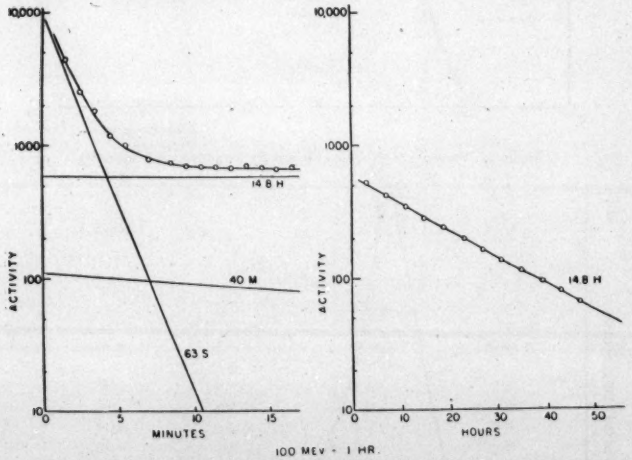


Figure 9. Decay curves obtained after bombardment of Mg with 100 MEV. x-rays.

These were obtained in the normal way, by irradiating specimens in the betatron beam, removing them, and then measuring as quickly as possible their β-activities with a suitable counter. The 40 m. activity was not identified. The 62·5 sec. activity is due to the ^{25}Na, reported by Huber *et al.* (1944), from the (γ, p) reaction. Absorption measurements of end-point energy, made by both groups of experiments, are in agreement (2·8 MEV.). The 14·8 hour activity (Figure 9) is apparently caused by ^{25}Mg (γ, p) ^{24}Na. Table 4 summarizes the results of Baldwin and Klaiber (1946).

Some probable cases of the (γ, αn) reaction with O and Na were also reported.

A Wilson chamber photograph from this paper is shown in Figure 10 (see Plate). Track 1 is a 6·2 MEV. α particle, and tracks 2 and 3 were identified as

protons. Track 4 is apparently the recoil nucleus (weak activities were observed which could be attributed to (γ, p2n) (γ, 3pn) and (γ, αn) or (γ, 2p3n) reactions). Figure 11 shows a similar star, and the tracks are probably due to the same kinds of particles as in Figure 10. Another interesting photograph (Figure 12), due also to Baldwin and Klaiber (1948), shows another form of nuclear evaporation (100 MEV. irradiation). The authors suggest that this might be either $^{12}C \rightarrow 2\alpha + 2p + 2n$ (threshold 35·4 MEV.) or, more likely, $^{14}N \rightarrow 3\alpha + p + n$ (threshold as in 20 MEV.).

On the latter assumption the quantum energy was found to be 41 MEV.

Baldwin and Klaiber (1948) tabulated the thresholds for various reactions Table 5.

Table 4

Element	Half-life	Activity	Probable isotope	Probable reaction
N	20·0 m.	weak	^{11}C 20·5 m.	^{14}N (γ, p2n) ^{11}C
Mg	62·5 s.	strong	^{25}Na 62 s.	^{26}Mg (γ, p) ^{25}Na
Mg	14·8 h.	strong	^{24}Na 14·8 h.	^{25}Mg (γ, p) ^{24}Na
Al	63 s.	moderate	^{25}Na 62 s.	^{27}Al (γ, 2p) ^{25}Na
Al	14·8 h.	moderate	^{24}Na 14·8 h.	^{27}Al (γ, pn) ^{24}Na
Si	2·5 m.	strong	^{28}Al 2·4 m.	^{29}Si (γ, p) ^{28}Al
Si	6·7 m.	strong	^{29}Al 6·7 m.	^{30}Si (γ, p) ^{29}Al
Si	15 h.	weak	^{24}Na 14·8 h.	^{28}Si (γ, 3pn) ^{24}Na
P	6·7 m.	moderate	^{29}Al 6·7 m.	^{31}P (γ, 2p) ^{29}Al
Pb	4·5 m.	strong	^{204}Tl 4·1 m.	^{206}Pb (γ, 2pn) ^{204}Tl

Table 5

Reaction	Threshold (MEV.)	Reaction	Threshold (MEV.)
^{12}C (γ, p) ^{11}B	15·9	^{12}C (γ, 2n) ^{10}C	32·3
^{12}C (γ, n) ^{11}C	18·6	^{12}C (γ, 2pn) ^{9}Be	33·6
^{12}C (γ, 2p) ^{10}Be	27·0	^{12}C (γ, 2α 2p 2n)	35·4
^{12}C (γ, pn) ^{10}B	27·2		

Ogle, Brown and Conklin (1947) reported in early 1947 that X-rays from a 20 MEV. betatron had given activities, with various target materials, that could be explained by the (γ, p) process

$$^{A}_{Z}A + \gamma \rightarrow {}^{A}_{Z-1}A + p,$$

although the fact that neutrons were produced in the betatron installation prevented the findings from being conclusive. The above authors then irradiated ^{9}Be since ^{10}Be (produced in the (n, γ) process) has a half-life of 10^5 years and so would cause no confusion. (γ, n) would give ^{8}Be, with the immediate production of two α-particles. The (γ, p) reaction would give ^{8}Li (β active, $E_{max} = 12$ MEV., half-life = 0·88 sec.).

The tests showed that the activity in ^{9}Be had a half-life of $0{\cdot}88 \pm 0{\cdot}03$ sec., with β-particles ranging in energy up to 12 MEV. A simple filter experiment eliminated confusion with the (n, α) ^{6}He β's that could have been produced, since their maximum energy would have been 3·7 MEV., although evidence for the production of ^{6}He was also obtained. The threshold of the Be (γ, p) reaction was found to be 18 ± 1 MEV. A later note by Conklin and Ogle (1947) reports the detection of the α-particles produced by the ^{8}Be into which the ^{8}Li from Be (γ, p) decays.

In the meanwhile, Hirzel and Wäffler (1947) published a detailed report of an investigation of the (γ, p) reaction in ten elements, using the γ-rays from ^{7}Li (p, γ) ^{8}Be.

To minimize the effect of background counts due to γ-radiations from the tube etc., the counter (Al cathode) used to determine the activity produced by the (γ, p) process was operated in anti-coincidence with a bank of ten counters, which were exposed to the background radiations. This procedure resulted in a 5- to 6-fold reduction of background count. The (γ, p) activities were equivalent to approximately 10–100 counts/min. and the reduced background to approximately 2–3/min.

The activities of the products of the (γ, p) reaction were compared with the (γ, n) activity produced in ^{63}Cu. Very careful precautions (such as the use of specially purified hydrogen in the ion source), were taken to reduce as far as possible the importance of the (n, p) reaction which might otherwise have swamped the (γ, p) activities. The production of neutrons from the Li target by D bombardment and from ^{7}Li (α, n) ^{10}B (the α's being produced from ^{8}Be formed in the (p, α) reaction with ^{7}Li) was considered. The final results (Tables 6 and 7) were thought to be sensibly free from such spurious effects.

Table 6. (γ, p) Cross Section compared with that for (γ, n) in ^{63}Cu

Target element	Radioactive product	Half-life	σ_{rel} (%)
^{25}Mg	^{24}Na	14·8 h.	2·83
^{26}Mg	^{25}Na	60 s.	1·56
^{29}Si	^{28}Al	2·3 m.	3·45
^{30}Si	^{29}Al	6·7 m.	1·26
^{50}Ti	^{49}Sc	57 m.	1·62
^{53}Cr	^{52}V	3·9 m.	8·1
^{77}Se	^{76}As	26·75 h.	4·8
^{98}Mo	^{97}Nb	75 m.	3·5
^{105}Pd	^{104}Rh	44 s. / 4·2 m.	7·3
^{111}Cd	^{110}Ag	24·5 s.	4·4
^{112}Cd	^{111}Ag	7·5 d.	5·3
^{113}Cd	^{112}Ag	3·2 h.	6·0
^{117}Sn	^{116}In	13 s.	2·9
^{118}Sn	^{117}In	117 m.	1·5

Table 7. Theoretical and Experimental Values of $\sigma(\gamma, p)/\sigma(\gamma, n)$

Reaction	Experimental	Theoretical (E_n=10 Mev.)	Theoretical (E_n=7·5 Mev.)
^{50}Ti (γ, p) ^{49}Sc	0·054	$1·2\times10^{-2}$	—
^{53}Cr (γ, p) ^{52}V	0·324	$4·7\times10^{-2}$	—
^{77}Se (γ, p) ^{76}As	0·048	$2·1\times10^{-4}$	$2·0\times10^{-3}$
^{98}Mo (γ, p) ^{97}Nb	0·028	$2·6\times10^{-4}$	$1·4\times10^{-3}$
^{105}Pd (γ, p) ^{104}Rh	0·055	$7·8\times10^{-5}$	$7·7\times10^{-4}$
^{111}Cd (γ, p) ^{110}Ag	0·034	$1·8\times10^{-5}$	$2·9\times10^{-4}$
^{112}Cd (γ, p) ^{111}Ag	0·040	$1·9\times10^{-4}$	$1·0\times10^{-3}$
^{113}Cd (γ, p) ^{112}Ag	0·046	$4·6\times10^{-5}$	$5·4\times10^{-4}$
^{117}Sn (γ, p) ^{116}In	0·022	$7·7\times10^{-6}$	$1·4\times10^{-4}$
^{118}Sn (γ, p) ^{117}In	0·011	$2·1\times10^{-5}$	$1·8\times10^{-4}$

In Table 7 the quotient (γ, p) cross section/(γ, n) cross section for the same element, derived from the (γ, n) cross section for ^{63}Cu, is compared with theoretical

values determined by the method of Weisskopf and Ewing (1940). Columns 3 and 4 involve the use of neutron binding energies, E_n, equal to 10 and 7·5 MeV. respectively.

It is readily seen that the values for $\sigma(\gamma, p)$ are much higher than would be predicted from the Weisskopf and Ewing theory. Further work on this problem is to be expected, and, indeed, Courant (1948) suggested, in order to explain the theoretical discrepancy, that (cf. Weisskopf and Ewing 1940) the incident γ-quantum may be absorbed by dipole interaction with a single proton (which would be immediately ejected) in the struck nucleus, instead of forming a compound nucleus with it. Courant reports that preliminary calculations with the dipole model give (γ, p) cross sections of the correct order of magnitude.

D. H. Wilkinson (Cavendish Laboratory) pointed out to us during 1947 that the (γ, p) reaction with silver, amongst other elements, might be important in the study of the photo-disintegration of D, using D-loaded photographic plates, since it might lead to the production of unwanted protons.

Klaiber, Luebke and Baldwin (1947), and Luebke, Klaiber and Baldwin (1947) have made detailed studies of the heavily ionizing particles ejected from nuclei by their 100 MeV. X-radiation. They identified most of the particles as protons, and proceeded to study with care the effect of multiple scattering, in the gas of the Wilson chamber, on the measurement and interpretation of the tracks. Analysis of the few tracks that gave mass values deviating a little from the proton mass showed in every case that the particle was probably a proton. The existence of quite large spurious curvatures, obtained in zero magnetic field, is shown by Figure 13 (see Plate), in which the track has a radius of curvature of 60 cm. (13 cm. range).

Perlman and Friedlander (1948) investigated in some detail various (γ, n), (γ, p), $(\gamma, 2n)$ and $(\gamma, 2p)$ reactions using X-rays emitted from a tungsten target under bombardment by 50 or 100 MeV. electrons.

The irradiations were monitored by measurement of ^{11}C or ^{18}F activities in polystyrene or LiF respectively, bombarded with each target investigated. A mica-window Geiger counter was used to measure activities, and the various corrections for window thicknesses, self-absorption in the sample etc. were carefully made. The X-ray spectra were not known, so that energy variations of cross sections could not be discussed. It is stated, however, that preliminary measurements on the X-ray beam energy distribution give at least approximate agreement with theoretical predictions that the number of quanta in a small energy interval is inversely proportional to the mean energy of the interval.

The results are summarized in Tables 8 and 9.

Table 8. Relative Yields of (γ, p) Reactions : $^{14}N(\gamma, n)^{13}N = 1{\cdot}00$ at each Energy

Parent isotope	Product isotope	Product half-life	Product betas and energies (MeV.)	Relative yield 100 MeV.	50 MeV.
$^{30}_{14}Si$	^{29}Al	6·8 m.	$\beta-2{\cdot}5$	5·8	6·6
$^{57}_{26}Fe$	^{56}Mn	2·62 h.	$\beta-0{\cdot}75$ (20%) 1·04 (30%) 2·81 (50%)	7·6	7·6
$^{62}_{28}Ni$	^{61}Co	1·74 h.	$\beta-1{\cdot}1$	5·4	5·0
$^{99}_{42}Mo$	^{97}Cb	76 m.	$\beta-1{\cdot}4$	5·0	3·1
$^{102}_{44}Ru$	^{101}Tc	14·5 m.	$\beta-1{\cdot}2$	3·7	3·6

Table 9. Relative Yields of (γ, 2n) and (γ, 2p) Reactions: Yield of $^{14}N(\gamma, n)^{13}N = 1{\cdot}00$ at each Energy

Reaction	Product half-life	Product betas and energies (MeV.)	Relative yield 100 MeV.	50 MeV.
^{12}C (γ, 2n) ^{10}C	not observed	$\beta+3{\cdot}4$	$<0{\cdot}003$	
^{19}F (γ, 2n) ^{17}F	72 s.	$\beta+2{\cdot}1$	0·22	0·15
^{31}P (γ, 2n) ^{29}P	not observed	$\beta+3{\cdot}6$	$<0{\cdot}1$	
^{63}Cu (γ, 2n) ^{61}Cu	3·2 h.	$\beta+0{\cdot}9$ (78%)	3·3	2·5
^{27}Al (γ, 2p) ^{25}Na	62 s.	$\beta-2{\cdot}7$ (45%), $\beta-3{\cdot}7$ (55%)	0·15	0·14
^{31}P (γ, 2p) ^{29}Al	6·7 m.	$\beta-2{\cdot}5$	0·20	0·15
^{63}Cu (γ, 2p) ^{61}Co	1·8 h.	$\beta-1{\cdot}1$		0·16

The tables are accompanied in the paper by lists of references.

Tests were made (Perlman and Friedlander 1948) to check that neutrons and protons produced by the X-rays in the target and surroundings did not seriously affect the measurements of the primary (e.g. (γ, n)) reactions. Attention is directed by the above authors to various interesting features of the results, viz. the sudden increase in (γ, n) yields in the neighbourhood of mass 60, and the high yield of ^{187}Re (γ, n). It is also pointed out that (γ, p) yields for the cases studied vary less than the (γ, n) yields for the same parent isotopes and are of the same order as (γ, n) yields for species below copper, and it is suggested that this is difficult to reconcile with a mechanism of evaporation from a compound nucleus.

3.4. *Energy Loss of Fast Electrons and Positrons and Pair Production*

The loss of energy which an electron may be expected to suffer due to radiation and ionization, in travelling through matter of a given atomic number, has been calculated by Heitler (1944), and the results for the total energy loss per centimetre for different initial energies are shown in Table 10 for air, water and lead. (Pair production processes will of course give extra absorption.)

Table 10. Energy Loss (MC^2/cm.) for Various Incident Energy E

E (MC^2)	0·01	0·1	1	10	100	1000	10000
Air	0·073	0·013	0·0042	0·0045	0·0080	0·035	0·30
Water	69·4	12·2	3·80	3·98	6·9	28	240
Lead	260	65	23	41	205	1950	19500

The measurement of the energy loss has been described by Turin and Crane (1937), Curtis (1938), Fowler and Oppenheimer (1938) and others. Thin laminae of aluminium, carbon or lead are generally used as absorbers in a Wilson cloud chamber, the electrons and positrons being produced in a separate target or in the cloud-chamber wall. The results may be summarized as follows: for absorbing materials of low atomic number (aluminium, carbon etc.) the energy losses agree reasonably well with theory, but for lead the values found by some authors (but not by others, Turin and Crane 1937) are much higher (~50% greater) than those to be expected (see, for example, Laslett and Hurst (1937)), although the extent of the errors appears to be generally greater for the higher energies. No difference is found between the energy losses for electrons and those for positrons of the same energy. Part at least of the discrepancy for lead may

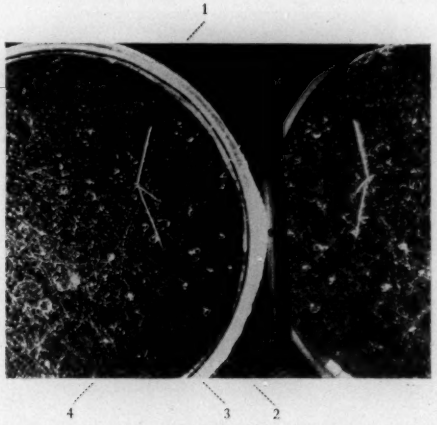


Figure 10.

Figure 11.

To face page 96

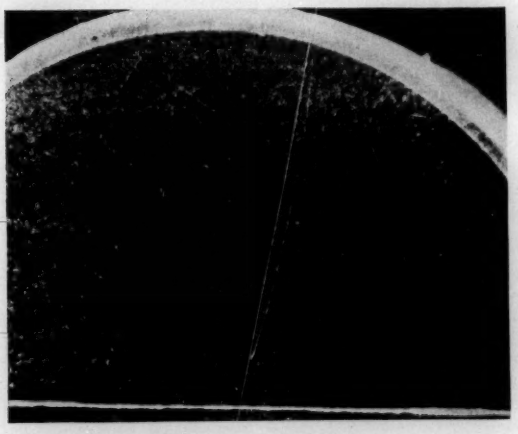


Figure 13.

Figure 12.

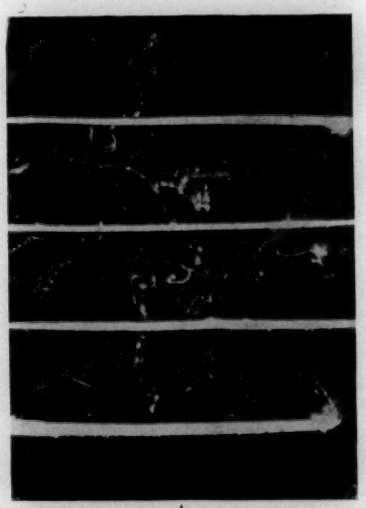


Figure 14.

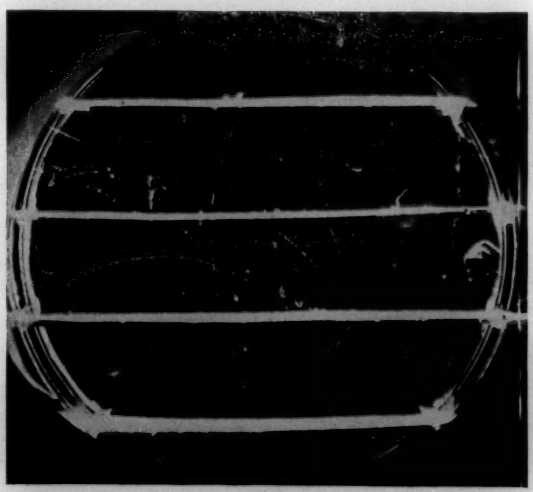


Figure 15.

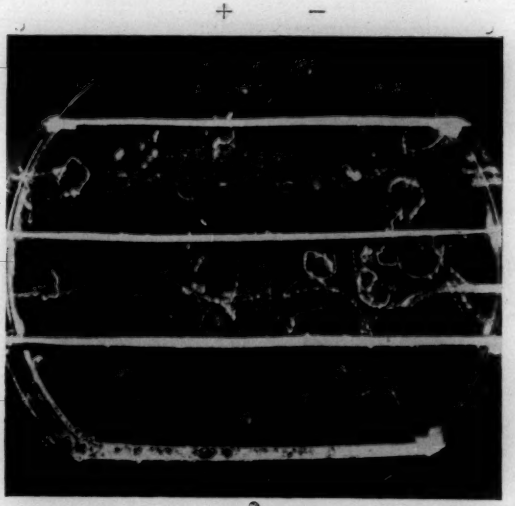


Figure 16.

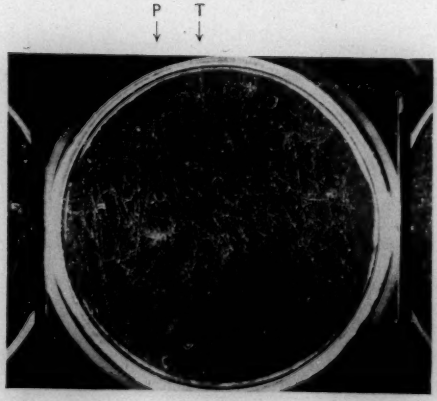


Figure 17.

be accounted for by the scattering of the electrons as they traverse the absorber, since this scattering increases the effective thickness of the absorber and becomes appreciable for high atomic numbers. Fowler and Oppenheimer (1938) found that the scattering (single scattering above $\theta = 13°$, multiple scattering below that angle) in lead for 5–17 Mev. electrons and positrons agrees with theory (Mott–Rutherford, etc.).

Preliminary results of measurements of the scattering of fast electrons produced by betatron X-rays have been described (Bosley *et al.* 1947), using (*a*) a single lead plate 0·013 in. thick as absorber, the electrons being produced in the cloud chamber wall, and (*b*) a set of four lead plates the first of which, 0·013 in. thick, was used as a target for the production of pairs by the X-rays, the electrons then passing through the other three plates of increasing thicknesses (0·010 in., 0·020 in., and 0·030 in.) suffering increasing energy losses.

A series of some 600 photographs (stereoscopic pairs) has been taken (Nash, Bosley and Craggs 1948). Figure 14 shows a triplet (a pair formed in the field of an electron) found in the course of this work, and Figure 15 shows a typical

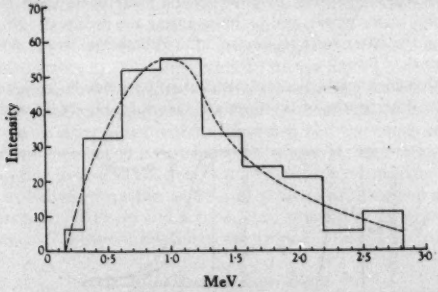


Figure 18. 2 Mev. betatron radiation spectrum deduced from analysis of Compton recoils.

electron track. The scattering is particularly marked in one of the lead plates. Figure 16 shows an electron/positron pair and apparently large angle scattering of the positron. These results were obtained with a 20 Mev. betatron. Figure 17, included by the courtesy of Dr. G. S. Klaiber (G.E. Laboratory, Schenectady), shows another triplet and a pair taken with the G.E. 100 Mev. betatron.

The above work on the production of electron/positron pairs enables also an investigation of the betatron spectrum to be made, following earlier work by many authors on the investigation of certain γ-ray spectra. Typical work in the latter field was carried out by Crane *et al.* (1935), Gaerttner and Crane (1937), Halpern and Crane (1939), Fowler *et al.* (1938) and Kruger and Ogle (1945). In comparing radiation intensities from pair energy distributions it is necessary to assume that proper allowance for variations of cross-section with energy can be made. Heitler (1944) gives full information on this subject and it is generally supposed, as a result of experimental work, that the treatment is valid at least for the range of energies covered in this Report.

Lasich and Riddiford (1947) published a spectrum (Figure 18) for their 2·8 Mev. betatron, determined from measurements on Compton recoil electrons

ejected from a 0·5 mm. carbon plate in a Wilson chamber. Their spectrum differs considerably from that due to a thin target. Figure 2 (Heitler 1944, p. 170) shows the results calculable for thin targets, and it is seen that the spectrum, roughly triangular at the lower energies, becomes more rectangular at the higher energies. (It will often be useful to replot Figure 2 to give numbers of quanta against energy).

Lasich and Riddiford attributed the departure of their spectrum from that due to a pure bremsstrahlung process to secondary emission of electrons from the injector and vacuum chamber and the existence of electron orbits with radii less than that of the equilibrium orbit.

The present authors (Bosley *et al.* 1948) have carried out preliminary experiments on a Kerst type 20 MeV. betatron and have used pair production, and photo-protons from the photo-nuclear disintegration of D, to give information on the beam radiation. Normalized at one point, the spectra obtained by the two widely different methods agree reasonably well. The pairs were produced, as described above, in lead plates in a Wilson chamber and the photo-protons were produced, in a simple scattering camera, by allowing the X-ray beam from the betatron to strike a "heavy wax" target, and by intercepting the protons in Ilford C2 plates. The tracks in the latter were measured, after processing, in the normal manner. The recent work of Powell and his school at Bristol has, of course, greatly increased the ease with which such measurements may be made in technical and other laboratories. The results of the tests are shown in Figure 19 and a discussion

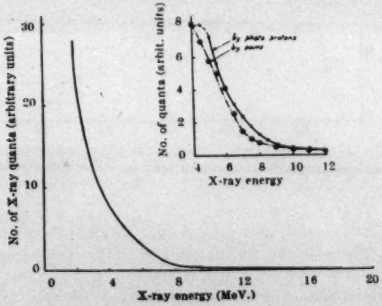


Figure 19. Experimentally determined 20 MeV. betatron radiation spectrum.

of them has been given elsewhere (Bosley *et al.* 1948). It is very likely that further work on betatron spectra, which is clearly urgent, will modify the conclusions of Lasich and Riddiford and of Bosley *et al.*

It was felt that Lasich and Riddiford's explanation of the difference between the measured spectra and those to be expected from thin targets could not hold in the present case (Bosley *et al.* 1948) because the vacuum chamber was too narrow to permit the existence of orbits of appreciably less energy than the equilibrium value at any time and because the X-ray pulses as detected by a Geiger-Müller counter were emitted always at the same time in the acceleration cycle—that is, the amount of scattered radiation was very small. It seemed, therefore, that the

electron beam was sensibly monochromatic and that the conversion process was not simply that which would occur in a thin target. It was shown qualitatively that the electrons, which would not be completely stopped in their first encounter with the target, could be expected to oscillate inside the vacuum chamber with an increasing amplitude and so eventually to bombard the inner wall of the chamber, thus producing a background of "extremely thick target" radiation which might be of the distribution found experimentally.

Pending the measurements of spectral energy distributions, various authors have published measurements of photo-nuclear thresholds and variations of the cross section of such processes with quantum energy on the assumption that the beam radiation gave a bremsstrahlung spectrum (Baldwin and Klaiber 1947, 1948). It is possible, in the authors' opinion, that for the 100 MeV. betatron the spectrum of the total emitted radiation is more likely to give a closer approach to a bremsstrahlung spectrum than is the total radiation from the 20 MeV. betatrons. However, even in the latter case it is to be expected that, if the normal (thin) targets were used, as in general practice, and efforts were made to exclude scattered radiation and to observe only radiation emitted at the target, then the resulting spectra would be closer to the bremsstrahlung type than those of Figures 18 and 19. Heitler (1944, pp. 161–177) discusses angular distributions, etc.

Baldwin and Klaiber (1947) have indicated the need for spectrum measurements with betatrons and have described a labour-saving modification of the electron pair method described above. Baldwin and Klaiber point out that X-ray spectra can be determined without observing both members of each pair. The energy of one member of a pair does not, of course, determine the energy of the quantum that produced the pair, but it is possible to calculate statistically the energy distribution of quanta which will produce an observed distribution of positron energies. Baldwin and Klaiber then proceed in detail to show how this may be done. No experimental data are given.

3.5. *Miscellaneous Experiments*

Laughlin, Skaggs, Hanson and Orlin (1947) have described, in a preliminary note, the results of experiments on electro-disintegrations (e, ne). A 20 MeV. betatron was modified to allow the electron beam to emerge from the vacuum chamber (a technique described by Skaggs, Almy, Kerst and Lanzl (1946)) and the beam was then allowed to strike various targets. The cross sections for the reactions with ^{63}Cu, ^{107}Ag and ^{109}Ag at 16 MeV. were stated to be 17, 49, and 77×10^{-29} cm^2 respectively, some 400 times less than the photo-disintegration cross sections.

Pollock and his collaborators (Elder, Langmuir and Pollock 1948) have observed intense visible radiation from the high energy electrons accelerated in a 70–80 MeV. synchrotron. The beam was first visible at about 30 MeV. as a dull red spot, which became very brilliant and of a bluish-white colour at 80 MeV. The spectral distribution of the radiation was measured and found to agree well with the results of calculations due to Schwinger.

ACKNOWLEDGMENTS

The authors wish to thank Dr. G. C. Baldwin and his colleagues in the G.E. Betatron Laboratory, Schenectady for their great kindness, shown in sending

prints for Figures 10, 11, 12 and 13, and for permitting reproduction of the hitherto unpublished photograph shown in Figure 17.

Acknowledgment is also due to the following bodies for permission to reproduce figures: *Physical Review*, Figures 1, 3–13; Oxford University Press, Figure 2; *Nature*, Figure 19; Messrs. Lasich and Riddiford and *Journal of Scientific Instruments*, Figure 18; Metropolitan-Vickers (Mr. B. G. Churcher and Mr. F. R. Perry), Figures 14–16.

REFERENCES

Baldwin, G. C., and Klaiber, G. S., 1946, *Phys. Rev.*, **70, 259**; 1947, *Ibid.*, **71,** 3; 1948, *Ibid.*, **73,** 1156.

Baldwin, G. C., Klaiber, G. S., and Hartzler, A. J., 1947, *Rev. Sci. Inst.*, **18,** 121.

Baldwin, G. C., and Koch, H. W., 1945, *Phys. Rev.*, **67,** 1.

Becker, R. A., Hanson, A. O., and Diven, B. C., 1947, *Phys. Rev.*, **71,** 466.

Bosley, W., 1946, *J. Sci. Instrum.*, **23,** 277.

Bosley, W., Craggs, J. D., and Nash, W. F., 1947, *Nature, Lond.*, **160,** 790.

Bosley, W., Craggs, J. D., Nash, W. F., and Payne, R. M., 1948, *Nature, Lond.*, **161,** 1022.

Bothe, W., and Gentner, W., 1939, *Z. Phys.*, **112,** 45.

Chadwick, J., and Goldhaber, G., 1934, *Nature, Lond.*, **134,** 237.

Collins, B. G., Waldman, B., and Guth, E., 1939, *Phys. Rev.*, **56,** 876.

Conklin, R., and Ogle, W., 1947, *Bull. Amer. Phys. Soc.*, **22,** No. 5, 8.

Corson, D. R., and Wilson, R. R., 1948, *Rev. Sci. Instrum.*, **19,** 207.

Courant, E. D., 1948, *Bull. Amer. Phys. Soc.*, **23,** No. 3, 30.

Crane, H. R., Delsasso, L. A., Fowler, W. A., and Lauritsen, C. C., 1935, *Phys. Rev.*, **48,** 125.

Curtis, B. R., 1938, *Phys. Rev.*, **53,** 986.

Dunworth, J. V., 1940, *Rev. Sci. Instrum.*, **11,** 167.

Elder, F. R., Gurewitsch, A. M., Langmuir, R. V., and Pollock, H. C., 1947, *J. Appl. Phys.*, **18,** 810.

Elder, F. R., Langmuir, R. V., and Pollock, H. C., 1948, *Phys. Rev.*, **74,** 52.

Fowler, W. A., Gaerttner, E. R., and Lauritsen, C. C., 1938, *Phys. Rev.*, **53,** 628.

Fowler, W. A., and Oppenheimer, J., 1938, *Phys. Rev.*, **54,** 320.

Fry, D. W., Gallop, J. W., Goward, F. K., and Dain, J., 1948, *Nature, Lond.*, **161,** 504.

Fry, D. W., R.-S.-Harvie, R. B., Mullett, L. B., and Walkinshaw, W., 1947, *Nature, Lond.*, **160,** 351.

Fry, D. W., and Walkinshaw, W., 1949, *Rep. Prog. Phys.*, **12,** 102 (London: Physical Society).

Gaerttner, E. R., and Crane, H. R., 1937, *Phys. Rev.*, **52,** 582.

Gibson, W. M., Green, L. L., and Livesey, D. L., 1947, *Nature, Lond.*, **160,** 534.

Gray, L. H., 1928, *Proc. Roy. Soc.* A, **122,** 648.

Halpern, J., and Crane, H. R., 1939, *Phys. Rev.*, **55,** 200.

Haxby, R. O., Shoupp, W. E., Stephens, W. E., and Wells, W. H., 1941, *Phys. Rev.*, **59,** 57.

Heitler, W., 1944, *The Quantum Theory of Radiation* (Oxford: University Press).

Heymann, F. F., and Bosley, W., 1949, in course of publication.

Heyn, F. A., Aten, A. H. W., and Bakker, C. J., 1939, *Nature, Lond.*, **143,** 516.

Hirzel, O., and Wäffler, H., 1947, *Helv. Phys. Acta*, **20,** 373.

Huber, O., Lienhard, O., Scherrer, P., and Wäffler, H., 1944 *Helv. Phys. Acta*, **17,** 139.

Kerst, D. W., 1942, *Rev. Sci. Instrum.*, **13,** 387; 1946, *Nature, Lond.*, **157,** 90.

Klaiber, G. S., Luebke, E. A., and Baldwin, G. C., 1947, *Phys. Rev.*, **71,** 649.

Koch, H. W., and Robinson, C. S., 1948, *Rev. Sci. Instrum.*, **19,** 36.

Kruger, P. G., and Ogle, W. E., 1945, *Phys. Rev.*, **67,** 273.

Lasich, W. B., and Riddiford, L., 1947, *J. Sci. Instrum.*, **24,** 177.

Laslett, L. J., and Hurst, D. G., 1937, *Phys. Rev.*, **52,** 1035.

Laughlin, J. S., Skaggs, L. S., Hanson, A. O., and Orlin, J. J., 1947, *Bull. Amer. Phys. Soc.*, **22,** No. 6, 11.

LAURENCE, G. C., 1937, *Can. J. Res.*, **15**, 67.
LAX, M., 1947, *Phys. Rev.*, **72**, 61.
LEWIS, W. B., 1942, *Electrical Counting* (Cambridge : University Press).
LUEBKE, E. A., KLAIBER, G. S., and BALDWIN, G. C., 1947, *Phys. Rev.*, **71**, 657.
MUEHLHAUSE, C. O., and FRIEDMAN, H., 1946, *Phys. Rev.*, **69**, 691.
MYERS, F. E., and VAN ATTA, L. C., 1942, *Phys. Rev.*, **61**, 19.
NASH, W. F., BOSLEY, W., and CRAGGS, J. D., 1948, in course of publication.
OGLE, W., BROWN, L., and CONKLIN, R., 1947, *Phys. Rev.*, **71**, 378.
PERLMAN, M. L., and FRIEDLANDER, G., 1948, *Phys. Rev.*, **74**, 442.
ROBERTS, R. B., MEYER, R. C., and HAFSTAD, L. R., 1939, *Phys. Rev.*, **55**, 417.
ROSSI, B., and GREISEN, K., 1941, *Rev. Mod. Phys.*, **13**, 240.
SKAGGS, L. S., ALMY, G. M., KERST, D. W., and LANZL, L. H., 1946, *Phys. Rev.*, **70**, 95.
STEVER, H. G., 1942, *Phys. Rev.*, **61**, 38.
TURIN, J. J., and CRANE, H. R., 1937, *Phys. Rev.*, **52**, 63, 610.
WEISSKOPF, V. F., and EWING, D. H., 1940, *Phys. Rev.*, **57**, 472.
WESTENDORP, W. F., and CHARLTON, E. E., 1945, *J. Appl. Phys.*, **16**, 581.
WIEDENBECK, M. L., 1946, *Phys. Rev.*, **69**, 235.
WIEDENBECK, M. L., and MARHOEFER, C. J., 1945, *Phys. Rev.*, **67**, 54.
WILSON, C. W., 1945, *Radium Therapy* (London : Chapman and Hall).
WILSON, R., COLLIE, C. H., and HALBAN, H., 1948, *Nature, Lond.*, **162**, 185.
YUKAWA, H., and SAKATA, S., 1937, *Sci. Pap. Inst. Phys. Chem. Res.* (Japan), **31**, 187.

LINEAR ACCELERATORS

By D. W. FRY and W. WALKINSHAW
Telecommunications Research Establishment, Great Malvern, Worcs.

CONTENTS

PAGE
§ 1. Introduction 102

Part I—Electron Accelerators
§ 2. Dynamics of particles in R.F. fields 105
2.1. Axial stability 105
2.2. Radial stability 107
§ 3. Theory of microwave resonators and waveguides 109
3.1. Basic forms of accelerator 109
3.2. The travelling wave accelerator 113
3.3. The standing wave accelerator 115
§ 4. Acceleration in the pre-relativistic region 118
§ 5. Methods of constructing accelerators 118
§ 6. Feeding accelerators with R.F. power 120
§ 7. Particle injection 124
§ 8. Performance characteristics 125
§ 9. Future development of electron accelerators 126

Part II—Heavy Particle Accelerators
§ 10. A standing-wave proton accelerator 128

§ 1. INTRODUCTION

The production of high energy particles or quanta is of importance in nuclear physics and in medicine for therapeutic and other purposes. Particles with energies up to about 50 Mev. are of interest in the medical field, and energies of 500–1,000 Mev. and beyond are required for nuclear research. It is not surprising therefore to find that since about 1930 considerable thought has been given to finding the most practical ways of accelerating particles.

The early methods produced the accelerating fields either at low frequency A.C. or at D.C.; Lauritsen and Bennett (1928) used a group of high voltage transformer windings in series, Cockcroft and Walton (1930) a number of high voltage vacuum rectifiers in series, and Van de Graaff (1931) a development of the principle of the electrostatic machine. Cockcroft and Walton (1932) were the first to produce artificial disintegrations, and since that time all three methods have been widely used for accelerating particles, particularly the electrostatic generator, which was reviewed in the *Reports on Progress in Physics*, Vol. XI (Van de Graaff, Trump and Buechner 1947). Accelerators built along these lines may be made to have several very useful properties, such as a very narrow energy spectrum and relatively high beam intensity, but the need for establishing an accelerating voltage equal to the desired particle energy between an insulated electrode and ground severely limits the useful energy range for which they may be used.

Several methods were soon suggested for overcoming this limitation. For example, in the cyclotron (Lawrence and Edlefson 1930) positive ions were made to travel in semi-circles of gradually increasing radius between the poles of a large D.C. magnet, and were accelerated by a relatively low A.C. voltage of fixed

radio frequency produced between two D-shaped electrodes placed between the magnet poles. The R.F. voltage between these dees is usually of the order 10,000–100,000 volts, and particles gain twice this energy for every revolution they make, so that after many revolutions they emerge with energies much greater than the peak value of the R.F. voltage used in the accelerating process. More recently the radio frequency applied to the dees has been modulated to allow positive ions to be accelerated into the relativistic region, and alpha particles have been accelerated to approximately 400 MeV. (Brobeck *et al.* 1947).

Another method accelerated particles by passing them through a series of cylinders, applying an alternating potential of the correct phase between them. This principle was used by Wideroe (1928) to accelerate potassium ions to 50 kv. with a maximum applied voltage of 25 kv. and was later developed by Sloan working first with Lawrence (Sloan and Lawrence 1931) and then with Coates (Sloan and Coates 1934) to accelerate mercury ions to 1·26 and 2·85 MeV. Beams and Snoddy (1933, 1934) also accelerated particles in a straight line using a rather different technique from Sloan and his co-workers. Instead of using electrodes connected to an R.F. source, their electrodes were connected to the appropriate points on a loaded transmission line to which an impulse voltage was applied. By arranging the velocity of propagation of the pulse along the line to be the same as the particle velocity through the electrode system the particles were accelerated due to the series of impulses which they received at the gaps between the electrodes. By this method protons were accelerated to several million volts.

Beams and Trotter (1934) have also used this method in modified form to accelerate electrons to 1·3 MeV. Because the velocity of the potential surge in a loaded transmission line is always less than the velocity of light, it was now necessary to apply the voltage to the series of accelerating electrodes from separate transmission lines of different lengths.

A further method of overcoming the high potential insulation difficulty is to use an induction accelerator or betatron. This method has been used very successfully by Kerst (1941) and many other workers to accelerate electrons. It was successfully demonstrated for the first time much later than the first demonstrations of the other methods, but electrons have already been accelerated in this way to energies as high as 100 MeV. (Westendorp and Charlton 1945).

More recently several new methods by which particles might be given very high energies have been proposed, and in some cases successfully tried out (Morrison 1947, Woodyard 1948). The most important of these is the synchrotron (McMillan 1945, Veksler 1945, Oliphant, Gooden and Hide 1947), an orbital machine in which the guiding magnetic field increases with time as in the betatron, but in which the energy is supplied to the particle by an R.F. resonator. This machine is suitable for electron acceleration with a constant R.F. applied to the resonator, but can also be used for proton acceleration by applying frequency modulation. Other cyclic methods include the microtron (Veksler 1945, Schwinger (see Schiff 1946)) and the cavitron (Post 1946).

This Report will review the progress which has been made in the last three to four years in the development of the linear accelerator. The major part of the work to be reviewed has been carried out on centimetric and decimetric wavelengths for the acceleration of electrons, and only one important project has been briefly described for accelerating heavy particles (protons); this used a wavelength of 1½ metres. Most of the statements made in the Report

are therefore only intended to apply at decimetric wavelengths and below. The design of heavy particle accelerators presents technical problems quite different from those of electron accelerators since, on account of their larger mass, the heavier particles travel at much slower velocities than the electrons for the same kinetic energy. These problems are discussed briefly in § 10.

Interest in particle linear accelerators was renewed towards the end of the second World War (1939–45), due to the many advances which were made in the development of very high frequency power sources, mean pulse powers of 0·1–2·0 MW. for several microseconds at repetition rates up to 500 p.p.s. for frequencies in the range 150–3,000 Mc/s. becoming available. These high pulse powers, together with their associated new circuit techniques, offered the possibility of linear acceleration of particles comparing favourably on technical and economic grounds with some of the other methods (Schiff 1946, Woodyard 1946). The main reasons are the low radiation loss by the particle in the relativistic region, and the ease with which relatively large mean currents might be injected and ejected. Further, the cost would be approximately a linear function of peak energy of the particles, whereas the cost of any form of cyclic accelerator is proportional to at least the square of the peak energy, if not to its cube. That these considerations are important when electrons are to be accelerated may be easily seen. For by the theory of energy radiation from a relativistic particle travelling round a circular orbit the loss of energy per revolution by radiation may be written to a first order approximation (cf. Page and Adams 1940) as

$$V_R = (6 \times 10^{-9}/R)(V/V_0)^4 \text{ electron volts,} \qquad \ldots\ldots(1)$$

where R = radius in metres, V_R = energy lost by radiation per revolution, V = total energy of particle in ev., V_0 = rest energy of particle in ev.

If electrons were given energies of 10^9 volts in any form of cyclic accelerator—for example, a synchrotron having an equilibrium orbit radius of three metres—then the loss of energy per revolution would be approximately 30 kev. This figure increases with the fourth power of the particle energy and therefore would become excessive for energies greater than 2,000–3,000 MeV. No such losses occur for a particle of constant velocity travelling in a straight line.

No satisfactory method of computing the maximum peak current which might be injected into a linear accelerator has been worked out, but experience gained so far may be used as a guide. Fry *et al.* (1948) have shown that in a 2-metre electron accelerator giving a peak energy of 4 MeV., 30% of the R.F. power fed into the accelerator could be supplied to the beam. On a long accelerator designed without allowance for beam loading it may not be practicable to supply as large a fraction of the available power as this to the electrons without reducing the peak energy achieved; under these conditions 10–15% of the power seems practicable. When beam loading is taken into account in designing the accelerator there seems no reason why the beam power should not be a large fraction of the total power input. It is shown in the body of this Report that effective shunt impedances of 30 megohms/metre should be attained on a long accelerator. Therefore for one designed to give a peak energy of 300 MeV. in 100 metres with the particles 35° ahead of the crest of the wave, the mean pulse power fed into the accelerator would have to be 45 MW. neglecting beam loading. If 10% of this power was fed into the beam the mean current in the pulse would be 15 ma., which for a duty cycle of 2,500 : 1 gives an average current of 6 μa.

It is found that the mean target currents which may be established in a betatron or synchrotron are substantially less than this. Blewitt (1946) and others have given expressions for the maximum charge which may be held in a synchrotron vacuum chamber when the mutual repulsion between particles is just balanced by the focusing force due to the guiding magnetic field. Taking into account the duty cycle of the machine, the mean target current corresponding to this charge may be written

$$I_{\text{mean}} = \frac{mc^2/e}{120} \cdot \beta^2 \left(\frac{\Delta}{r_0}\right)^2 \cdot \frac{2\pi r_0 f}{c} \cdot \frac{1}{(1-\beta^2)} \text{ amp.}, \qquad \ldots\ldots(2)$$

where mc^2/e = rest + kinetic energy of electrons at injection (in ev.), β = ratio of injection velocity to that of light, Δ = mean radius of vacuum chamber, r_0 = equilibrium orbit radius, f = frequency of magnetic cycle. For a 300 MeV. synchrotron with an injection voltage of 100 kv., $\Delta = 7{\cdot}5$ cm., $r_0 = 125$ cm., and $f = 5$ c/s.; this corresponds to a mean current of 1·0 microampere. In practice it has been found on lower energy betatrons, e.g. 20 MeV., that the mean circulating currents are never more than one-tenth the value predicted theoretically and are often about one-hundredth of it. Thus the mean currents of a betatron or synchrotron are one-sixtieth to one-six-hundredth of those to be expected from a linear accelerator.

For reasons such as these, and the possibility that the cost of producing linear accelerators might be appreciably less than that for other forms of accelerator, work has been undertaken in a number of places to appraise more fully the possibilities and limitations of the linear acceleration of particles in the energy region of a few MeV. up to 1,000 MeV. and even higher. The most serious limitation at present is the lack of suitable valves with pulsed powers which are sufficiently large to reduce the total number required to a reasonable figure.

This Report will be written in two parts. The first part will review recent advances in linear accelerators, with particular application to the acceleration of electrons. The second part will deal in a similar way with heavy particle accelerators, although the published work to be reviewed in this field is very small compared with that to be covered by the first part.

PART I—ELECTRON ACCELERATORS

§ 2. DYNAMICS OF PARTICLES IN R.F. FIELDS

2.1. *Axial Stability*

The basic aim of all linear accelerators is to produce a progressive wave of controlled phase velocity less than or equal to the velocity of light with a component of electric field in the direction of propagation of the wave. Particles injected in the correct phase into a wave of this type will gain energy from the wave and form stable bunches. A means by which this suitably controlled type of wave may be produced is therefore required. It is one of the characteristics of the circular smooth-walled waveguide, discussed in many modern text-books, that waves of this type may be propagated along them but with phase velocities greater than the velocity of light. Several satisfactory methods of slowing down the wave have been suggested. All of these produce, in addition to the desired wave, other components which have, in general, still slower phase velocities and some of which are reflected in the opposite direction. In an accelerator these waves, since they are travelling at appreciably different velocities from the particles,

cannot contribute to the mean particle energy. They can therefore be ignored in a first analysis of the particle dynamics. To this approximation the problems of phase and radial stability of particles in a linear accelerator have been studied by Ginzton, Hansen and Kennedy (1948), Slater (1948a), R.-S.-Harvie (1948a) and Walkinshaw (1948).

Ignoring for the moment the possibility of radial motion, it will be seen that particles can be accelerated in this way by reference to Figure 1. This represents the variation in amplitude of the axial electric field at a given instant in time. The wave is progressing from left to right with an increasing phase velocity such that a positive particle at B (D etc.) is maintained in a constant accelerating field represented by the horizontal broken line. The phase velocity of the wave is then the same as that of the particle and its dependence on distance z along the axis is

$$V = Ez \sin \phi_s + \text{constant}, \qquad \ldots\ldots(3)$$

where V is the total energy (rest and kinetic) of the particles of charge e, E is

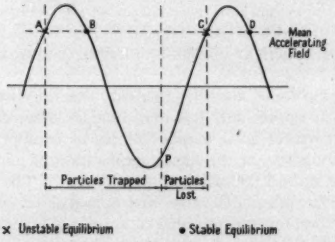


Figure 1.

the peak field, and ϕ_s the phase of the particle relative to zero field. Particles displaced from B (or D) find themselves in a field which produces either too much acceleration or too little, thus restoring them to the phase stable point B (or D). Likewise particles at A (or C) are in unstable equilibrium.

Slater (1948a) has examined stable phase oscillations such as this by an application of the well known W.K.B. approximation of quantum mechanics (e.g. Slater and Frank 1933). The axial equation of motion of the particle in the progressive wave can be expressed as

$$\frac{d}{dt}(m\dot{z}) = eE \sin\left(\omega t - \int k\,dz\right), \qquad \ldots\ldots(4)$$

where m is the relativistic mass of the particle, e the charge, k the phase constant varying with distance z, E the amplitude of the field, and ω the angular frequency.

By referring the motion to the stable phase position, the first order equation gives the following solution for the frequency of the phase oscillation :

$$\frac{\omega_0}{\omega} = \left[\frac{E\lambda \cos \phi_s}{2\pi\beta} \frac{V_0^2}{V_s^3}\right]^{\frac{1}{2}}, \qquad \ldots\ldots(5)$$

the amplitude of which is inversely proportional to

$$[V_s^2 \beta_s^3 E \cos \phi_s]^{\frac{1}{2}} \qquad \ldots\ldots(6)$$

where ϕ_s is the stable phase position, V_s is the total energy (rest and kinetic) of the stable particle, β_s is the ratio of its velocity to that of light, V_0 is the rest energy, λ is the free-space wavelength of the R.F. wave of angular frequency ω. Only qualitative trust can be placed in the above expressions, since the W.K.B. approximation holds only for slow rates of change which do not always apply for the conditions being considered. However, it will be seen that there is damping of the amplitude of the phase oscillations with increasing particle energy both for slow and relativistic speeds, and that the phase oscillations have generally a frequency much lower than that of the wave, almost ceasing to exist in the relativistic region. These points are of extreme importance in accelerating electrons. Referring again to Figure 1, if particles are injected continuously into the wave at its initial phase velocity, only those in the region A to, say, P will be trapped in resonance with the wave and bunch around B. Those starting with too large an initial displacement, from P to C in advance of the stable phase, fall back beyond A and are lost. The trapping region decreases with increasing mean accelerating field and ceases to exist when this equals the peak field. In the initial stages of an electron accelerator some compromise must be made between the rate of acceleration and the fraction of electrons trapped. Once relativistic velocities have been reached, however, the phase oscillations are so slow that it is possible to bring the bunched particles to the peak of the wave without serious debunching. Ginzton *et al.* (1948) find that the reduction in electron velocity relative to the wave at high energies is roughly in inverse proportion to the cube of the distance traversed along the accelerator. This rapid decrease in relative motion between particle and wave suggests that, whilst for particles in the non-relativistic region it is necessary to increase the phase velocity with distance, at some stage in the relativistic region it is sufficient to have a uniform phase velocity equal to that of light. Slater (1948 a) and Walkinshaw (1948) have examined an accurate analytical solution for the special case of electron motion in a wave of uniform phase velocity. For wave velocities less than that of light, the electron motion relative to the wave resembles that of a circular pendulum, being either in resonance with the wave in oscillatory motion around the stable phase position (now at zero field) or continuously moving either forward or backward if its initial velocity is much different from that of the wave. Of particular interest is the degenerate solution for the case of wave velocity equal to that of light. There can now be no oscillatory motion since the electrons cannot move forward relative to a wave of this velocity. For a range of initial phases, and provided the electron energy is 'sufficiently large', the particle will continue to gain energy indefinitely from the wave. If the electrons are not to drift more than $\pm 10°$ from the peak of the wave 'sufficiently large' means greater than $(\pi V_0^2/2E\lambda \cos 80°) = (2{\cdot}3/E\lambda)$ MV., where $E\lambda$ and V_0 are in megavolts.

2.2. *Radial Stability*

The unique solution of Maxwell's electromagnetic field equations for an axially symmetrical travelling wave, found in many text-books (e.g. Stratton 1941), may be written

$$\left.\begin{aligned} E_z &= -E_0 J_0(\chi\rho)\sin(hz-\omega t) \\ E_\rho &= (h/\chi)E_0 J_1(\chi\rho)\cos(hz-\omega t) \\ Z_0 H_\phi &= (k/\chi)E_0 J_1(\chi\rho)\cos(hz-\omega t) \end{aligned}\right\} \quad \ldots\ldots(7)$$

where z, ρ and ϕ are the usual cylindrical coordinates, χ and h are respectively the radial and axial components of the phase constant, k, with $k^2=\chi^2+h^2$, and Z_0 is the intrinsic impedance of free space. Ginzton, Hansen and Kennedy (1948), Slater (1948a) and R.-S.-Harvie (1948a) have all shown, using those equations, that the radial force on a particle is $e(E_\rho-\dot{z}H_\phi/c)$, which may be rewritten as

$$F=(h/\chi)eE_0J_1(\chi\rho)\cos(hz-\omega t)(1-v\dot{z}/c^2). \qquad \ldots\ldots(8)$$

In the region where phase stability exists this force is defocusing, although it is to be noted that as the particle velocity approaches that of light, the radial focusing magnetic force becomes almost equal to the electric force and the defocusing action is only slight.

R.-S.-Harvie (1948a) points out, however, that radial stability of the particle can be obtained by the application of a paraxial magnetic field, the magnitude of which has to be greater than

$$\frac{1}{300}\left[\frac{4V_0^2E_z\cos\phi_s}{V_s}\frac{\pi}{\lambda}\frac{c}{\dot{z}}\right]^{\frac{1}{2}} \text{ oersteds}, \qquad \ldots\ldots(9)$$

where V_0 is the rest energy of the particle, V_s is its total energy in electron volts, E_z is the peak accelerating field in v/cm., λ is the R.F. wavelength in cm., and ϕ_s is the phase of the particle with respect to the zero field.

Another method which has been suggested for maintaining stability resembles that of the Sloan and Lawrence ion accelerator. This depends on the higher order wave components which were neglected in the preceding analysis. These fields have the effect of localizing the radial force at the entrances and exits of the accelerating gaps which set up the R.F. field. A particle arriving at the centre of the gap when the field is at its maximum spends a greater portion of its time in the first half of its transit where the field is focusing than in the second where it is defocusing. There is therefore a very small region (Mullett 1946a) of stable phase positions near the peak of the travelling wave where both radial and phase stability exist. For electrons the tolerance required of the amplitude of the axial field to maintain this phase is likely to make the method unreliable for producing stability.

Schultz *et al.* (1947) claim that a theoretical investigation indicates that both axial and phase stability can be obtained in a series of cavity resonators which are mutually uncoupled. This is not in agreement with Mullett's conclusion and appears to ignore the effect of the radial electric impulse which the particle receives on entering and leaving each cavity.

Still other methods of focusing which have been proposed depend on distorting the field in the accelerating gaps by using foils (Alvarez 1946) or grids (Mullett 1946a, Gabor 1947a), although these methods do not appear to have been used in electron accelerators. At relativistic energies the focusing problem becomes less serious and it is possible to dispense with the focusing magnetic field. Slater (1948a) has pointed out that whilst the radial momentum of the particle remains constant in the absence of a radial field, the mass of the particle increases with its energy, with the result that the rate of spread decreases. The increase in radius is given by

$$\delta\rho=\phi_0\frac{V_1}{E_z}\ln\left(\frac{V_2}{V_1}\right), \qquad \ldots\ldots(10)$$

where V_1 and V_2 are respectively the initial and final electron energies, E_s is the accelerating field, and ϕ_0 is the initial angle which the particles make with the beam. For example, if electrons are injected with an energy of 2 MeV. into an accelerating field of 10 MV/metre, with an initial angular beam spread $\phi_0 = 10^{-3}$ radians, then the increase in radius will only be 1·24 mm. at 1,000 MeV. For electron acceleration it may be concluded that particle stability can be achieved provided the beam is well collimated by the time the particles reach relativistic velocities.

§ 3. THEORY OF MICROWAVE RESONATORS AND WAVEGUIDES

3.1. *Basic Forms of Accelerator*

The main basic forms which the R.F. system of a linear accelerator may take are shown in Figures (2*a*) to 2(*g*). The first shows a circular waveguide with metal irises, in which the iris holes provide for the passage of both electron beam and radio-frequency power. We shall see that, by a proper choice of guide dimensions, wave patterns of the required type with phase velocities less than that of light can be propagated along it. Such a guide has been used with a matched termination by Fry *et al.* (1947), Ginzton *et al.* (1948) and Starr (1948) for electron acceleration.

Alternatively the ends of this guide may be closed, setting up a standing wave instead of a travelling wave pattern. This type of excitation has been favoured by Halpern *et al.* (1946), Haxby *et al.* (1946), Slater (1948a), Newbery and Willshaw (1948) and Lawton and Hahn (1948).

Single cavity structures have been successfully operated by Bowen, Pulley and Gooden * (1946), Mills * (1948), Hereford † (1947) and Snoddy and Beams ‡ (1948), and alternative methods of coupling several cavities have been described by Schultz *et al.* § (1947), Allen and Symonds || (1947), Mullett ¶ (1946b), Cullen and Greig ** (1948) and Hudspeth †† (1946).

In addition to the above structures on which experimental work has been done there were a number of earlier suggestions by various workers. Starr (1945),

* Single re-entrant cylindrical cavity operated at 1,200 Mc/s. Mills' work was an extension of that of Bowen, Pulley and Gooden. One of its most interesting features is the production of a mean beam current of 70 microamperes in a focal spot size of approximately 0·5 mm. diameter (Figure 2 (*b*)).

† The accelerating system is a single re-entrant cavity resonator operating at 400 Mc/s. Double transit of the gap has been successfully achieved by reversing the beam in a magnetic field, but the measured energy spectrum is very wide. This appears to be an inherent limitation of the method.

‡ Single re-entrant cylindrical cavity operated at 400 Mc/s.

§ A series of cavity resonators is excited, each by a separate power amplifier at a frequency of 587 Mc/s., and is mutually uncoupled. Frequency and phase coherence are obtained by a master oscillator and the phase of the fields in each cavity is adjusted independently by a phase shifter. The use of power amplifiers overcomes the frequency stability problems associated with self-excited oscillators.

|| This system consists of three re-entrant cavities coupled at their outer surface, as shown in Figure 2 (*d*), and operated at 3,000 Mc/s. in the π-mode. The short gap across the cavity decreases the transit time of the particle considerably and improves the cavity efficiency. This type of design compares favourably with other resonant methods.

¶ A rectangular waveguide is loaded to obtain the required phasing with inductive corrugations (Figure 2 (*e*)), which are between one-half and three-quarters wavelength in depth. The particles would pass through holes in the corrugations at the maxima of the electric field. This system should be comparable in efficiency with an iris-loaded circular guide, but would be difficult to manufacture.

** A rectangular waveguide is folded in its broad face as shown in Figure 2 (*f*), and the electron beam is passed through the accelerator as indicated. In order to achieve a phase change of 180°, between sections along the electron path, the depth of each fold is an effective half-wavelength. This system is more difficult to manufacture than an iris-loaded guide.

†† The phase of the field in this case is delayed by coiling a rectangular guide in a spiral as shown in Figure 2 (*g*). Manufacturing and design difficulties make the system compare unfavourably with others of simpler construction.

described a number of structures of varying complexity derived from a concentric transmission line in which the inner conductor was broken into a number of quarter-wave sleeves providing an accelerating field along a central channel of the conductor. R.-S.-Harvie (1945) also suggested feeding a number of collinear half-wave resonators from a concentric line. Willshaw and Lamont (1945) considered loading a circular guide with a dielectric sheath. The properties of a dielectric sheath have also been studied by Frankel (1947), Bruck and Wicker (1947), and Oliner (1948). Most of these early suggestions have now been superseded by more efficient and less complicated structures.

Considerable attention has been given by all linear accelerator designers to determining the best type of waveguide or resonator system. It has been shown by Hansen and Richtmyer (1939) that for a single cavity having an axial electric field the geometrical form which gives the highest shunt resistance and, therefore, the greatest axial field for a given amount of power is a re-entrant cone as in Figure 2(*b*) with 9° cones. Akeley (1946a) on the other hand has produced a theoretically designed surface which will minimize the unwanted secondary field components. Difficulty in manufacturing and coupling optimum systems of this sort has forced most designers to adopt the more simple geometrical forms and in particular the cylindrical one with plane ends shown in Figure 2(*a*).

Much information, both theoretical and experimental, is available for this latter type of structure, the theoretical investigations by independent workers overlapping to a large extent. An upper limit to the energy which a relativistic particle can obtain from such a system is given by Ginzton *et al.* (1948) and Walkinshaw (1947) as occurring when there are between three and four resonators per R.F. wavelength. This assumes that the resonators oscillate in such a way that the particle passes each resonator centre when the accelerating field is a maximum, that equal power is fed into each resonator and that the size of the coupling holes between resonators is negligibly small. This is used as the standard of reference when discussing effective shunt impedance later in the Report.

The effective shunt impedance per unit length, i.e. the square of the average voltage gained by the particle per unit length for unit power dissipated in this length, is

$$3{\cdot}3\times10^{4}\frac{(\sigma/\lambda)^{\frac{1}{2}}}{n+2{\cdot}62}\left[\frac{\sin\pi/n}{\pi/n}\right]^{2}\text{ ohms per metre,}\qquad\ldots\ldots(11)$$

where n is the number of resonators per R.F. wavelength λ, and σ is the conductivity of the metal ($5{\cdot}8\times10^{7}$ mhos/metre for copper). For $n=3{\cdot}5$ and $\lambda=0{\cdot}1$ metre this gives approximately 100 megohms per metre. This means that if the optimum conditions could be fulfilled it would be necessary to feed 1 MW. mean pulse power of R.F. into an accelerator 100 metres long to produce electrons of 100 MeV. energy, or 100 MW. mean pulse power into the same length to produce 1,000 MeV. electrons. The effective shunt impedances which can be achieved in practice are always less than the optimum. It is of interest to note that if the coupling hole is negligibly small compared with the R.F. wavelength, equation (11) shows that the effective shunt impedance increases with decreasing wavelength. For a finite size of coupling hole there is, however, an optimum wavelength. Calculations show this wavelength to be 9 cm. and 17 cm. for hole diameters of 2 cm. and 4 cm. respectively with four cavities per free-space wavelength.

For an excellent qualitative description of the characteristics of the coupled row of resonators reference should be made to Slater's paper (1948a). Regarding

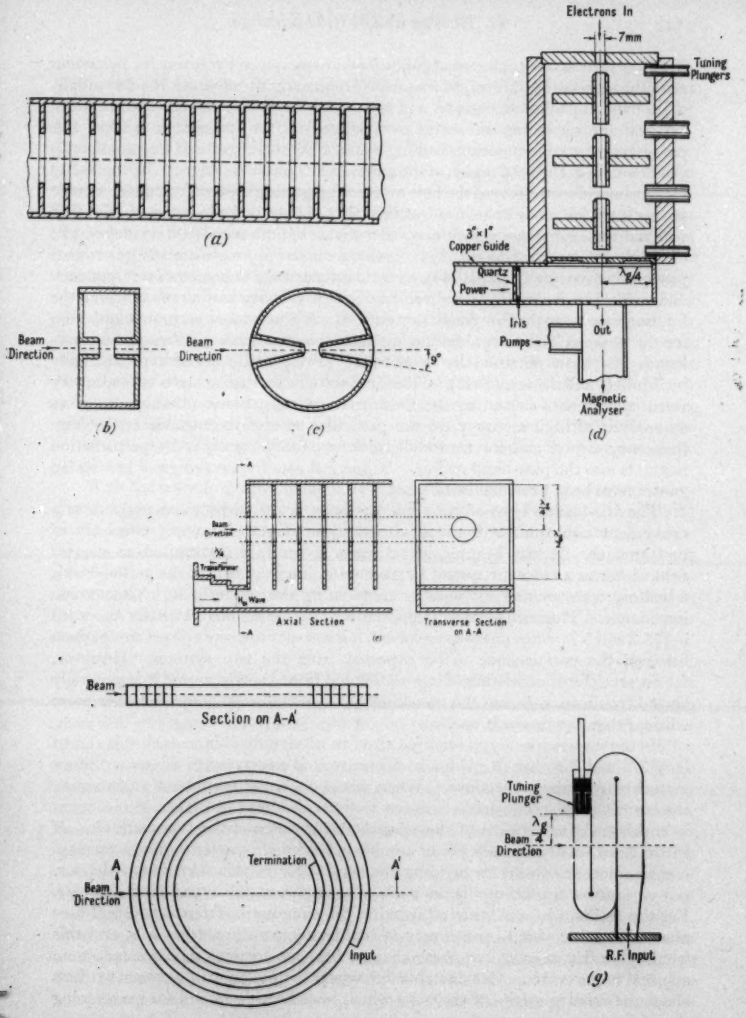


Figure 2.

(*a*) Axial section of iris-loaded circular waveguide.
(*b*) Re-entrant cylindrical cavity.
(*c*) Re-entrant spherical cavity.
(*d*) Re-entrant cavity structure of Allen and Symonds. (Taken from *Proc. Phys. Soc.*, 1947, **59**, 622.)
(*e*) Corrugated rectangular guide (Mullett).
(*f*) Spiralled waveguide–rectangular section (Hudspeth).
(*g*) Folded waveguide–rectangular section (Cullen and Greig). (Taken from *J. Appl. Phys.*, Jan. 1948.)

the structure as a circular waveguide loaded reactively by irises, its behaviour resembles a pass-band filter, having a low frequency cut-off which is substantially that of the unloaded waveguide, and also a high frequency cut-off when adjacent cavities are oscillating in antiphase (π-mode). For frequencies within this pass-band, a predominantly travelling wave of the form required by equation (7) can be set up in the axial region of the guide and its phase velocity can be controlled by varying either the size of the hole or the outer guide diameter. A characteristic of this wave for phase velocities less than that of light is the increase in axial field with radius. For phase velocities equal to that of light the axial field is independent of radius. Chu and Hansen (1947), using a similar approximate analysis to that previously given by Cutler (1944), have investigated the dependence on frequency and guide parameters of phase velocity, group velocity and attenuation for the limiting case near the low frequency cut-off. A much more accurate expression for the phase velocity in guides having more than five irises per free-space wavelength has been obtained by Walkinshaw (1948) and checked experimentally by Mullett and Loach (1948). Account is taken in that analysis of secondary wave components caused by the finite iris spacing. Slater (1948a) also gives an analysis of high accuracy for the particular case of the π-mode at the high frequency cut-off and for zero wall thickness, and extends it by perturbation methods into the pass-band region. Numerical data for the design of iris-loaded guides have been provided in all cases.

The iris-loaded type of structure has been accepted by most projects as a convenient compromise between simplicity of manufacture and efficiency of performance. It may be used in two ways to produce the desired accelerator field, either as a resonant system by short-circuiting the end of the guide or as a travelling wave system by suitably terminating the guide in its characteristic impedance. The main factors of importance in these methods are briefly discussed in §§ 3.2 and 3.3; since many are involved, it is not easy to make a direct comparison between the performance to be expected from the two systems. However, for an accelerator consisting of one section fed from a single source it is generally agreed by all workers that the travelling wave method may be made rather more efficient than the resonant one.

As the maximum energy required from an accelerator is increased, it is shown in §§ 3.2 and 3.3 that to maintain efficiency it is necessary to construct either system of a number of sections. When this is done, the problem of maintaining the correct phase relationship between sections also has to be taken into account in making any assessment of the relative performance of the two methods. If the sections are driven from power amplifiers fed from a master source, no serious complication arises due to dividing the accelerator in this way, but if the R.F. power sources are self-oscillators such as magnetrons the situation is different. For the oscillators, even when adjusted to the same nominal frequency, will vary about this value, due to inaccuracy of tuning, temperature drifts etc., and this would give rise to excessive phase changes between sections unless means were adopted to prevent it. Considerable differences of opinion exist between workers about the relative merits of the ways which have been proposed for overcoming this difficulty.

It has been claimed for the resonant method that it offers an important advantage in this respect compared with the travelling wave system, since the cavities may be used to stabilize the frequency of the oscillators. To assess the importance

of this it is necessary to examine carefully the physical processes involved in stabilizing the oscillator frequency. Willshaw (1947a, b), and Newbery (1948a) have considered this and have shown that the frequency near the beginning of the pulse is determined by the frequency of the unloaded oscillator, the series resistance and the length of the coupling line between them, whereas in the steady state, after the power in the resonator has built up, the frequency of the source is controlled by the resonator and a high degree of stability is ensured. The build-up time for the system is therefore a function of the difference between the resonant frequency of the cavity and the oscillating frequency near the beginning of the pulse. When these two frequencies are the same the build-up time is a minimum. In order to attain this it is necessary for the initial frequency to be within the resonance frequency band of the resonator. An expression for this build-up time has been given by Slater (1948a). In practice it is unlikely that the frequency tolerances of a number of magnetrons will be as narrow as this, and the build-up time may therefore be expected to be appreciably longer than the minimum value. No estimate of build-up time as a function of frequency tolerance of the source for a resonant system has been published, so that it does not seem possible at present to make an accurate comparison of its performance with that of a frequency-limited travelling wave system in order to assess the importance of this point.

With the travelling wave method, to maintain the necessary phase relationship between sections when driven from self-oscillators, it would be essential to provide sufficient coupling between the oscillators to synchronize them together in phase. Up to four magnetrons (Cowhig 1947) feeding into dummy loads have been locked in phase in this way, and it appeared that this number could have been increased. However, the use of a large number of self-oscillators all locked in phase is relatively complex, and it is desirable to develop higher power sources and to use fewer of them.

3.2. *The Travelling Wave Accelerator*

In the travelling wave accelerator, Figure 3, power is fed into one end of a matched corrugated waveguide and is dissipated in the walls as it propagates along the guide, the accelerating field being attenuated exponentially. When the diameter of the coupling holes in the irises is small, the attenuation is large since only a small fraction of the power entering a cavity is passed on to the next one. As the diameter of the coupling hole is increased the fraction of the power passed from one cavity to the next is also increased, the useful length of the waveguide becoming larger. In practice the size of the coupling hole which can be used should be determined by the diameter necessary to allow free passage of the electron beam. When this is fixed there is an upper limit to the length of waveguide system which may be used without loss of efficiency due to attenuation. It has been shown that this length is that in which approximately 90% of the power is dissipated in the guide walls, and 10% passes out of the end of the waveguide into the matching load. R.-S.-Harvie (1948a) and Ginzton *et al.* (1948) give extensive quantitative data showing effective shunt impedance as a function of length for differing iris apertures, when the outer diameter of the corrugations is adjusted to give a phase velocity equal to that of light. For a frequency of 3,000 Mc/s. and iris apertures of 2, 4 and 6 cm. the attenuation limited lengths are 3·5, 40 and 130 metres respectively when five cavities per wavelength and

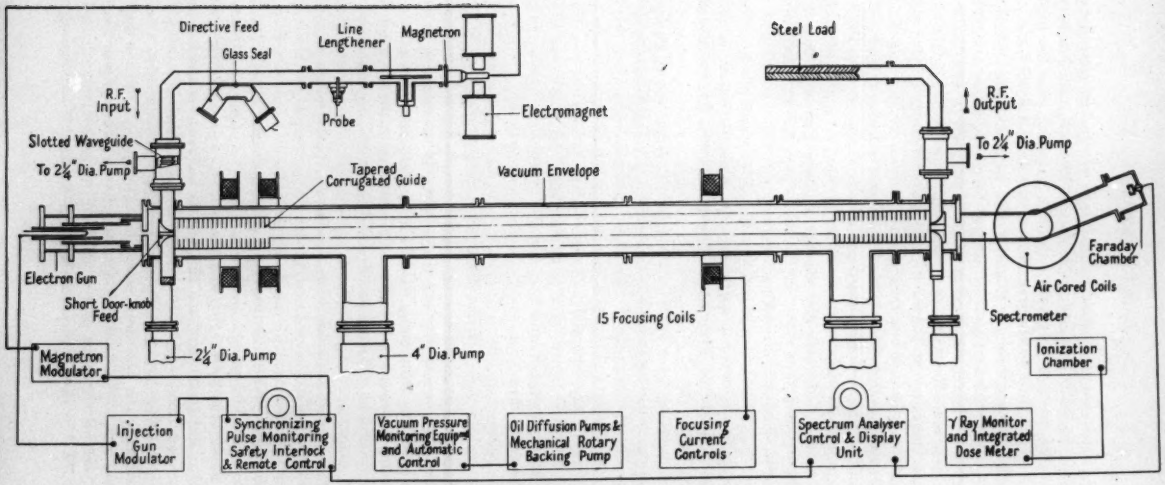


Figure 3. Schematic diagram of 4 MeV. linear electron accelerator.

theoretical ohmic loss in the guide are assumed. This corresponds to effective shunt impedances of 180, 1,400 and 3,000 megohms respectively. It should be noted that although this shows an increase in overall impedance with increasing optimum length, the average per unit length decreases. It is clearly advantageous to make a long accelerator in a number of sections using the smallest practical iris-hole diameter with its corresponding optimum section length rather than to make it in one long length.

For a variety of reasons it is often not possible to make the waveguide section as long as the attenuation limited length without loss of efficiency. This will occur if it becomes impossible to keep the electrons in an accelerating phase of the wave along the length of the accelerator. Such a phase error could occur if the frequency of the R.F. source was changed by a small amount, since it is a characteristic of a heavily loaded guide that it has a low group velocity (or high dispersion) due to the length of time required for the field in the cavities to build up. Therefore a small change in the radio-frequency makes a large change in the guide wavelength, giving a wave of incorrect phase velocity. The phase change $d\psi$, of a particle of velocity $\boldsymbol{c}$, in the wave at position z of the accelerator is given by R.-S.-Harvie (1948 a) as

$$d\psi = 2\pi \frac{z}{\lambda}\left[\frac{\boldsymbol{c}}{U} - 1\right]\frac{df}{f}, \qquad \ldots\ldots(12)$$

where df/f is the fractional change in frequency and $U/\boldsymbol{c}$ is the ratio of the group velocity of the wave to the velocity of light.

Typical estimated characteristics for travelling wave accelerators showing the relative importance of attenuation and frequency stability upon accelerator length and effective shunt impedance are given in Table 1. These figures are

Table 1. Estimated Characteristics for Travelling Wave Accelerator. Wavelength 10·0 cm.

Fractional inner radius a/λ		0·3	0·2	0·15	0·1
Attenuation limited characteristics	optimum length (m.)	63	19	6	1·7
	effective shunt impedance (MΩ/m.)	13	19	23	28
Frequency limited characteristics *	optimum length (m.)	12·5	4·4	2·5	0·85
	effective shunt impedance (MΩ/m.)	9·6	15	19	23·5

* Frequency stability ±0·01%.

calculated for electrons accelerated at the crest of the wave in a waveguide having five irises per wavelength with twice theoretical attenuation and a maximum permissible phase error of $2\pi/10$ radians between the electrons and the wave.

Computations of shunt impedance taking into account attenuation due to power losses in the guide walls and group velocity are involved and are not reducible to simple formulae. One result which is simple and useful relates the axial accelerating field E_0 to the power flow, W, along the guide for a phase velocity equal to that of light. This is, in M.K.S. units,

$$E_0 = (480W)^{\frac{1}{2}}(\lambda/\pi a^2), \qquad \ldots\ldots(13)$$

where a is the radius of the iris aperture and λ the R.F. wavelength.

3.3. *The Standing Wave Accelerator*

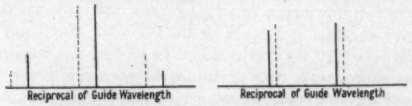
By closing the ends of a disc-loaded guide it assumes the characteristic of a resonator. Neglecting the power fed into the beam, all that fed into the resonant section is dissipated in its walls, none passing out of the end into a dummy load.

For this structure it is necessary to consider its properties during the time when the R.F. field is building up and when it reaches the steady state. Any resonant system requires $Q/2\pi$ cycles for oscillations in it to build up to within $1/e$ of their final amplitude. In any practical system the time width of the R.F. pulse should be sufficiently long for the oscillations to build up to at least 90% of their final value (Willshaw 1947b). Slater (1948a) has shown this to be comparable with the time of build-up of an attenuation limited travelling wave accelerator. In the steady state the attention of power distribution along a section is small. In addition it will have discrete modes of oscillation for those frequencies within its pass-band corresponding to an integral number of half guide wavelengths per guide length. Guide attenuation and frequency stability of the source are then no longer the factors which determine the optimum length of a section, as for a travelling wave accelerator; they become replaced by the frequency mode spectrum of the section. A useful method of calculating this for a multiple cavity resonator has been given by Willshaw (1946) based upon a circuit with equivalent lumped reactances at regular intervals. When the resonant section is fed from a self-oscillator the frequency separation required between adjacent frequency modes imposes a severe limitation on the maximum usable section length. For as a resonant section is lengthened, although its pass-band remains the same, the number of resonances within this band increase directly as the number of half wavelengths in its length. The spacing between these resonant frequencies is not uniform, but is greatest in the middle of the pass-band and least at its edges. Lawton and Hahn (1948) preferred to work their resonant accelerator in the middle of the frequency pass-band rather than at the high frequency cut-off of the band (π-mode), in order to take advantage of the greater mode separation in that region. In their accelerator they used four cavities per wavelength with a phase change of approximately $\pi/2$ between adjacent cavities, where π-mode operation occurs for two cavities per guide wavelength. Besides giving a greater mode separation, this reduces the effect of electron transit time across a cavity, so that the total acceleration per cavity for a given peak field is greater than for the π-mode of operation.

A similar result has already been noted in § 3.1 for collinear cavities fed with equal power, where three to four cavities per wavelength gave the largest effective shunt impedance. For an attenuation limited end-fed travelling wave accelerator the efficiency was 80% of that with uniform distribution of power along its length, neglecting the further reduction on account of the finite coupling holes. For a resonant accelerator, on the other hand, the mean power distribution along its length will be sinusoidal and the accelerator efficiency is now 50% of that with uniform distribution of power. This occurs because of the reflected wave which absorbs half the total power fed into the resonators without contributing to the mean energy of the particles.

Although there is a maximum efficiency for three to four cavities per wavelength for both travelling wave and resonant guides, there is a further important point to be noted regarding the π-mode. For a travelling wave accelerator the field can be analysed into component waves as described in § 3.1, with fundamental and other components of smaller amplitude as shown in Figure 4(*a*). If the accelerator is made resonant then the number of harmonics is doubled on account of the reflected waves (shown dotted in Figure 4(*a*)). For a resonant mode near the π-mode the space harmonics approach each other in pairs as shown in

Figure 4(*b*). There are now four components of almost equal amplitude and phase velocity, only one of which travels in the same direction and has exactly the same velocity as the particle. When the π-mode is ultimately reached the pairs of components coalesce and there is a consequent rapid increase in efficiency, attaining a value 70% of the optimum with three to four cavities. The dependence of shunt impedance on cavity number is shown in Figure 5 for both a resonant and a non-resonant section. The upper curve is calculated from equation (11) for a wavelength of 10 cm., and the lower curve is taken as half of this value because of the reflected wave. Because of its larger shunt impedance Slater (1948a) and many other workers have preferred to use the π-mode of oscillation



(*a*) Amplitude of harmonic components for frequency in centre of pass-band.

(*b*) Amplitude of harmonic components for frequency close to high frequency cut-off (π-mode).

Figure 4.

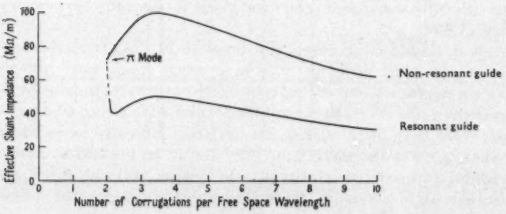


Figure 5.

although the maximum permissible section length for a specified mode separation is shorter than it is for the $\pi/2$ mode. It has been found practical by Slater (1948b), Lawton and Hahn (1948) and Newbery and Willshaw (1948) for this mode separation to be 0·1% without instability occurring.

In Table 2 estimated characteristics for standing wave sections of accelerators

Table 2. Estimated Characteristics for Standing Wave Accelerators. Wavelength 10·0 cm.

Fractional inner radius a/λ		0·4	0·3	0·2	0·1
Characteristics for 0·5% mode separation	optimum length (m.)	0·95	0·7	0·45	0·2
	shunt impedance (MΩ/m.)	4·0	5·0	8·0	11·5
Characteristics for 0·1% mode separation	optimum length (m.)	2·1	1·6	1·0	0·45
	shunt impedance (MΩ/m.)	4·0	6·0	9·0	13·0

are given for comparison with the estimated figures for travelling wave sections of Table 1.

Less quantitative information has been published about standing wave accelerators than about travelling wave ones; thus the estimated characteristics

of Table 2 may be less reliable and it is included with this reservation. The assumptions made were: (*a*) the ohmic loss was taken as twice theoretical, as is usually the case in practice; (*b*) 25% of the magnetron power was assumed to be dissipated in a resistive load to stabilize the oscillator; (*c*) the theory depends on an assumed field distribution at a radius equal to the iris aperture obtained from quasi-static considerations, and may introduce an error of ten to twenty per cent in the shunt impedances.

For a more detailed comparison of single short sections of travelling wave and resonant accelerators a paper by Newbery (1948 a) should be consulted.

§ 4. ACCELERATION IN THE PRE-RELATIVISTIC REGION

Except in the first few metres of an accelerator, the electrons travel essentially with the velocity of light, and much of the previous discussion has dealt with this latter uniform part of the accelerator. In § 2.1 of this Report, we described the particle motion in a wave with continuously increasing phase velocity leading to particle bunching at a stable phase position. T.R.E. (Fry *et al.* 1947) and Stanford (Ginzton *et al.* 1948) have adopted this method of attaining relativistic velocities. In the former case both the iris hole and the guide diameter have been varied in order that the accelerating field should remain constant along the accelerator and the stable phase position varied as desired. Details of the theoretical design together with computed tables for guide dimensions have been given by Walkinshaw (1948).

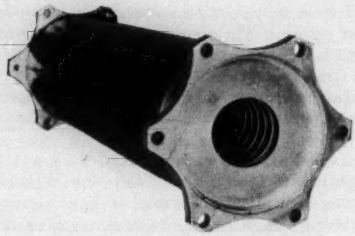
Although it should be theoretically possible to design a resonant type of accelerator on the same principle, it is in practice much more difficult unless the cavities are decoupled and fed independently, otherwise dimensional changes in one part of the guide affect the magnitude of the field in all other parts. Lawton and Hahn (1948) have attempted to get over this difficulty by subdividing the non-relativistic part of the accelerator into shorter lengths with uniform phase velocities which approach the velocity of light in steps, and with appropriate phase change between each section. This arrangement is complicated, however, and unlikely to provide a narrow energy spectrum.

Willshaw (1948) on the other hand has suggested using only one or two initial cavities with sufficiently high peak fields to attain relativistic velocities.

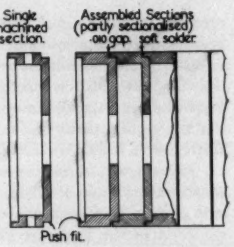
§ 5. METHODS OF CONSTRUCTING ACCELERATORS

In § 3, Figures 2(*a*) to 2(*g*), several geometrical forms which may be used for the waveguide cavity system are shown. With the exception of that shown in Figure 2(*c*) they are not ones which will give to a particle the maximum possible acceleration per cavity, but the loss in acceleration is not large and the other forms shown have a great advantage when the accelerator length becomes long, in that simplification in construction is possible.

Of great importance in the construction of these guides are the special methods of fabrication used in order to obtain electrical characteristics approaching the theoretical figures. Three of these methods are illustrated in Figures 6(*a*), 6(*b*) and 6(*c*). Figure 6(*a*) shows a length of a travelling wave accelerator made up of separate sections machined to be a push fit together, and then soft soldered on the outside Mullett and Loach (1948). Figure 6(*b*) shows a length of guide of a similar type, but this time the irises are formed of precision machined discs which are shrunk into accurately determined positions inside a honed



40 cm. length of corrugated circular guide.



Construction of corrugated circular waveguide. The sections are clamped together and the 0·100 in. gap is filled with soft solder as shown.

Figure 6 (*a*). Method of guide construction used by Fry, Harvie, Mullett and Walkinshaw (1947).

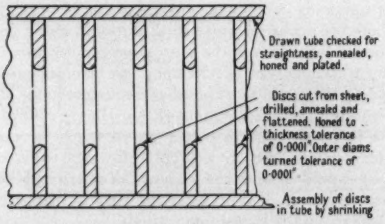
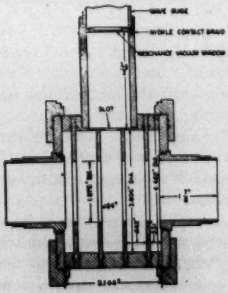

Drawn tube checked for straightness, annealed, honed and plated.
Discs cut from sheet, drilled, annealed and flattened. Honed to thickness tolerance of 0·0001". Outer diams. turned tolerance of 0·0001".
Assembly of discs in tube by shrinking

Figure 6 (*b*). Method of guide construction used by Ginzton, Hansen and Kennedy (1948).



Two-cavity, cavity-terminated, $\beta=0{\cdot}75$ accelerator design: $K=17{\cdot}75\%$, $\lambda_0=10{\cdot}705$ cm.

Two-cavity, $\pi/2$ mode design accelerator: $\beta=0{\cdot}75$, λ_0; 10·701 cm., $K=17{\cdot}75\%$

Figure 6 (*c*). Method of guide construction used by Lawton (1948). (From *J. Appl. Phys.*, June, 1948.)

cylinder (Ginzton *et al.* 1948). Figure 6(*c*) shows a construction which has been adopted for a resonant type of system. The cavities are formed by joining together alternate thin copper washers and large ring spacers at regular intervals (Lawton 1948). For the first and third of these methods it is found that the measured attenuation losses are approximately twice the theoretical figure. For the second method it is claimed that the measured attenuation approaches very close to the theoretical figures.

The manufacturing tolerances which have to be maintained must also be considered. First of all it is important to note that manufacturing tolerances only present a severe problem on long accelerators. For most of the accelerators recorded in the literature this problem has therefore not been a serious one. For long accelerators the manufacturing tolerances are determined by the necessity for the waves to remain in phase with the electrons within very close limits for the whole of their travel along the accelerator. If this is made up in a number of sections then it is important to adjust the phase correctly in each section to ensure that no cumulative phase error along the system occurs. These sections, whether they are of the travelling or standing wave types, are very sensitive to changes in their operating conditions, and departures in phase velocity from that required will therefore occur for quite small errors in guide dimensions, as well as small variations of frequency of the source and temperature changes along the waveguide. In a travelling wave accelerator phase errors occurring along sections may be corrected by the use of short phase correcting sections of slightly different phase velocity between the main sections (R.-S.-Harvie 1948a). For a standing wave accelerator any correction in phase required may be made by providing a fine tuning control in each resonant section. However, to avoid the considerable complexity which would arise if the number of correcting sections of tuning controls becomes very large, it is necessary to work to manufacturing tolerances of the order 0·0005 in. on the waveguide sections.

§ 6. FEEDING ACCELERATORS WITH R.F. POWER

The total energy gain V of an accelerator is proportional to $(PL)^{\frac{1}{2}}$ where P is the total power input and L is the total length. V may therefore be made large both by increasing P and by increasing L. It is difficult to find a sound reason upon which to decide the proper relative proportions of P and L, although Slater (1948a) has put forward the economic one that their costs should be equal. In practice it is found that for practical values of L the power P required soon exceeds the peak power available from the largest oscillators or amplifiers. The problems in the feeding of power into an accelerator therefore are those of providing the power required with the right frequency and phase and of distributing this power approximately evenly along the length of the accelerator.

Now it has been shown in § 3 that in a long accelerator a greater overall impedance can be obtained by subdividing it into sections, using their optimum iris-hole size rather than making it in one continuous length. The problems may therefore be divided into those which arise in feeding one section of an accelerator and those arising when the accelerator is made in more than one section.

A typical arrangement for feeding power from a single source into one section of a travelling wave accelerator is shown in Figure 3. The corrugated guide system is terminated in a suitable dummy load to avoid reflection of any unused power back into the magnetron. The input impedance characteristic of the

waveguide will then be purely resistive at the operating frequency of the system and for adjacent frequencies will not vary rapidly. Satisfactory methods of feeding power from standard rectangular waveguides into this form of impedance have been worked out by Mullett and Loach (1948) and Ginzton *et al.* (1948). When the optimum length for any chosen iris-hole diameter is used, some form of frequency discriminator and servo mechanism control of the frequency will be necessary, for a frequency stability of one part in 10^4 or better will be required. However, on accelerators which have been constructed up to the present the optimum length has not been used. No frequency control has therefore been provided since frequency stability was not a limiting factor. Work in this field remains to be done as frequency limited accelerators are built.

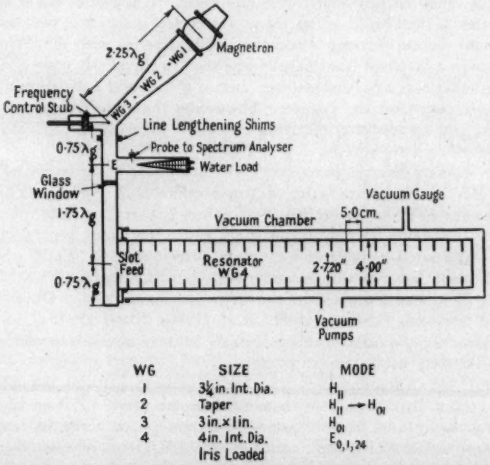


Figure 7. Method for feeding one section of a standing wave accelerator (Newbery and Willshaw). (Taken from *Nature*, 1948, **161**, 519.)

It will be required in the future to feed power at a higher level into a single section than can be obtained from any single source at present available. In principle this might be done either by paralleling a number of existing valves or by generating the power in larger units, and work has been started along both of these lines. The largest power source which has been used so far for linear accelerators is a magnetron operating on 10 cm. wavelength with 2 μsec. pulses having a mean power in the pulse of 2 MW. and a maximum pulse repetition rate of 500 per second. These magnetrons have been run in pairs, locked in frequency and phase (Cowhig and Jones 1947), delivering the sum of their powers into a dummy load, and there seems no reason why considerably more than two should not be run in parallel. However, such a system would be elaborate to set up and adjust. A more satisfactory solution to the production of higher peak powers

would be the design and development of larger individual power units, and work aimed at doing this is in progress. Devons (1947) is developing a magnetron on 25 cm. designed to give 30 μsec. pulses with a mean power in the pulse of 100 MW. Hansen (1948), working upon another possible solution of the problem, is designing a klystron amplifier to operate at 10 cm. wavelength giving 2 μsec. pulses also having a mean power in the pulse of 100 MW. Less spectacular developments in progress in the magnetron field aim at 10–20 MW. at 10 cm. wavelength from one valve. Even this would help materially with the development of electron linear accelerators.

To feed a section of a standing wave accelerator a different method has to be adopted. The resonator requires $Q/2\pi$ cycles to build up to within $1/\epsilon$ of its final value, and during that time its impedance as seen by the valve changes over a wide range starting from a very low value. If the valve is a self-oscillator, in order that it shall build up on the preferred frequency it is necessary to feed the resonant section through a resistive load in series with it. When this is done through a matched transmission line then in the steady state condition the resonator behaves as a parallel resonant circuit at the valve and acts as a frequency stabilizing element for the system. The cavity therefore determines the radio frequency, and no separate frequency control is necessary in the way that it is for the travelling wave system.

Many workers have investigated both theoretically and experimentally the problem of feeding resonant cavity electron accelerators. The circuit of Newbery and Willshaw (1948) for doing this is shown in Figure 7. They found that for a 24-cavity resonator having a Q of 9,000 and a frequency separation of 0·2% between the desired π-mode and the next nearest one, 65% of the magnetron power could be fed into the resonator section without difficulty arising. The remaining 35% was dissipated in the series stabilizing load. Other workers—Allen and Symonds (1947), Halpern *et al.* (1946), Slater (1948a), and Lawton (1948)—have also reported feeding 50–60% of their available power into resonators satisfactorily using similar circuits.

Slater (1948a) demonstrates that higher powers may be fed into a single resonant section by driving it with more than one valve. For an E_{010} multiple cavity such as he uses it is immaterial where it is fed along its length. His circuit is one section of Figure 8, which uses four HK7 magnetrons operating on a free-space wavelength of 10·0 cm. with a mean pulse power of 1 MW. for 3·4 μsec. Each resonant section is 4 ft. long, made up of 24 cavities, with one tuning plunger and a small pick-up loop in each section. To ensure the magnetrons oscillate in phase one is prepulsed by 0·1 μsec. and feeds a controlling signal into the cavity. The frequencies of the oscillations in the section are monitored by viewing the envelope of the rectified output from the pick-up loop. If the four magnetrons are not all locked to the same cavity mode the envelope pattern is filled in. Small adjustments are then made to the individual magnetron tuning controls until the envelope becomes a clean trace on the oscilloscope. The tuning controls of the four magnetrons are then ganged together and their tuning is adjusted to the frequency of the desired mode to give maximum signal into the cavity. Stable operation with a mode separation of 0·1% in frequency (but not 0·05%) between the wanted π-mode and its nearest neighbour has been found quite practical. When locked in frequency and phase it is found that the magnetrons will remain so for long periods.

To feed an accelerator made up of a number of optimum length sections the power to be supplied to all the sections must be synchronous in frequency and phase. At the power levels required it will be necessary to obtain this from several valves even when the larger power sources at present under development become available. The problem is, therefore, that of feeding from separate sources completely decoupled sections of an accelerator. How this is to be done will depend upon whether the power sources to be used are self-oscillators or power amplifiers.

For the standing wave accelerator Slater (1948 b) has produced one solution of the problem using self-oscillators. His complete circuit is shown in Figure 8. The accelerator is made up of five sections each fed by four magnetrons as described

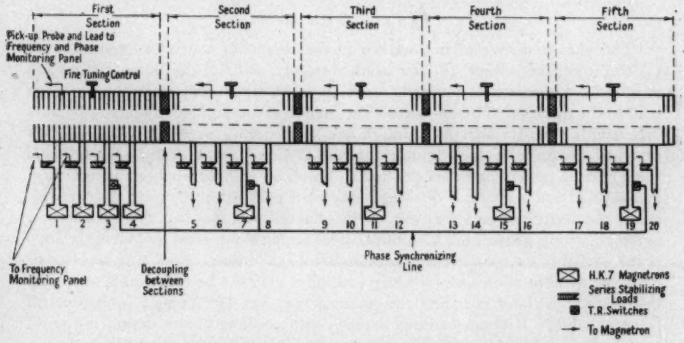

First Section
Second Section
Third Section
Fourth Section
Fifth Section
Pick-up Probe and Lead to Frequency and Phase Monitoring Panel
Fine Tuning Control
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
To Frequency Monitoring Panel
Decoupling between Sections
Phase Synchronizing Line
H.K.7 Magnetrons
Series Stabilizing Loads
T.R.Switches
To Magnetron

Figure 8. Schematic diagram of M.I.T. standing wave accelerator for 28 MeV. electrons. To ensure synchronization magnetron No. 11 is prepulsed 0·2 μsec. and magnetrons 3, 7, 15, and 19 0·1 μsec. before the remainder.
(We are indebted to Prof. J. C. Slater for very kindly supplying the information about this project.)

above. The frequency and phase locking is obtained in the following way. One of the magnetrons is prepulsed by 0·2 μsec. A small amount of power from the series stabilizing load of this magnetron is fed through T.R. switches to the magnetrons on the other four sections, which are prepulsed by 0·1 μsec. This provides them with the necessary signal to phase them correctly when they start oscillating. When these four magnetrons have built up, their respective T.R. switches fire preventing wastage of power along the synchronizing line. Each resonant section may therefore be separately tuned as described above and will be locked to start in the same phase. The remaining step in the tuning-up process is to compare the signals from different sections by feeding them into a crystal rectifier and viewing the rectified envelope as before. Any frequency difference may be eliminated by an adjustment of the fine tuning plunger mounted in each section.

An alternative method of feeding decoupled resonant sections using power amplifiers has been described by Schultz *et al.* (1947). The operating frequency used was 587 Mc/s. since triode power amplifiers delivering up to 600 kw. mean power in the pulse could be obtained. Two uncoupled E_{010} cavities were used

for the accelerating sections, each one being fed by a separate power amplifier having a power gain of 6–7, the two amplifiers being driven from a single master oscillator.

The principle of using power amplifiers capable of delivering large power in a pulse has the definite advantage that the frequency stabilization is no longer a problem since all the amplifiers may be driven from a well-established low power source. The 100 MW. pulsed klystron being developed by Hansen (1948) is of added importance for feeding long accelerators for this reason. Several alternative methods of using the high power sources under development for feeding accelerators made up of a number of sections have been proposed, but no work is possible until the sources themselves become available.

§ 7. PARTICLE INJECTION

The characteristics of importance in the design of the injector system are (*a*) the injection voltage, (*b*) the beam intensity, and (*c*) the beam collimation required. For much of the early work on accelerators little attention has been paid to the question of beam intensity so long as it was sufficiently large to obtain measurable output currents through the accelerating system. However, for practical use it is of considerable interest to know the maximum permissible beam current which can be accelerated. This limit will occur when the reduction in axial field due to the beam-loading reduces the peak energy given to the electrons below the desired value. Fry *et al.* (1948) have shown that in a short accelerator in which strong phase-bunching forces were acting along most of its length, 30% of the available R.F. power could be supplied to the beam. In a long accelerator designed without allowance for beam loading it may not be practicable to supply this amount without reducing the peak energy, but 15% would probably still be reasonable. If beam loading is taken into account in the design it should be practicable to feed a much higher percentage of the input power into the beam.

Beam collimation and the injection voltage to be used are, to some extent, inter-related problems. Hereford (1947) and Bowen, Pulley and Gooden (1946) have shown that an uncollimated beam of electrons with zero velocity may be used to inject into single cavity resonant accelerators. However, no figures are given to show the fraction of electrons leaving the filament which is accelerated. At the other extreme Slater plans to inject at 2 million volts into a standing wave accelerator from a pulsed Van de Graaff generator. The electron source of this electrostatic generator will be pulsed by photocell control through a small window in the side of its pressure chamber. The divergence of this beam will be of the order 10^{-3} radians. In § 2.2 it was shown that the subsequent spreading in such a beam might be expected not to be serious even on accelerators designed to give 1,000 MEV. electrons. The choice of injection voltage to be used in any accelerator must be made taking into account the relative complexity and reliability of the pulsed D.C. injecting source as compared with these factors for the initial section of an accelerator when it is designed for low voltage injection. A number of workers have found injection in the region 50–75 KV. to be a reasonable compromise. Fry *et al.* (1947) used a triode gun with a tungsten filament, Lawton and Hahn (1948) and Ginzton *et al.* (1948) used guns designed for use on x-ray equipment. For travelling wave accelerators injection voltages in this region provide the reasonable compromise, the manufacturing difficulties in corrugated guide construction increasing rapidly for phase velocities less than about 0·4 c.

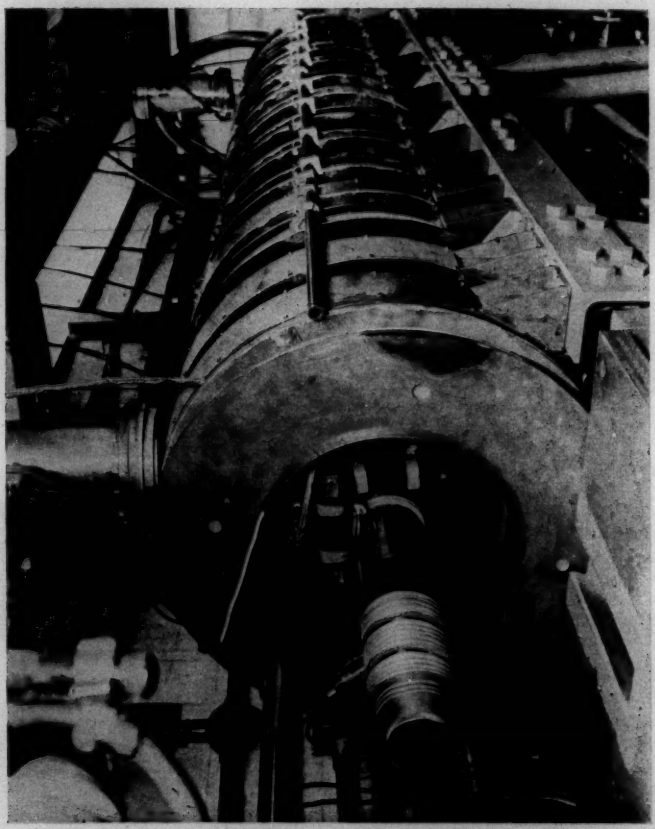


Figure 9.

A 4-MeV. linear electron accelerator showing injection gun.

To face page 125

§ 8. PERFORMANCE CHARACTERISTICS

In the previous sections of this Report it has been shown that a number of rather difficult technical problems have had to be solved if the possible technical advantages of the linear accelerator compared with the orbital accelerators were to be exploited. It is not surprising therefore that the majority of the experimental work done up to the present has been primarily directed towards finding the answers to some of these problems on low energy machines rather than to the production of high energies. The principal physical characteristics of the machines which have been built will now be considered, paying particular attention to (*a*) the peak energy at maximum intensity of the beam of particles, (*b*) the energy spectrum of the particles, (*c*) the maximum beam current together with the duty cycle of operation, and (*d*) the degree of collimation of the beam. It is also of interest to know the efficiency of accelerators measured as (Peak Energy)^2/(Total Power × Total Length) expressed in megohms/metre. These units are the same as those used for the guide effective shunt impedance per unit length in § 3.1, but they are now used to measure the overall efficiency of an accelerator and not the waveguide efficiency as they were previously.

Information about the accelerators already built, as far as it is available, has been summarized in Table 3, and a photograph of a 4 MeV. travelling wave accelerator is given in Figure 9. In Table 3 it is shown that electrons have been accelerated to energies in the region 0·3–6·5 MeV., and there can be little doubt that the accelerator which has been built by Slater will be made to give electrons with peak energies close to 28 MeV.

Most workers have measured their peak energies by rather crude magnetic analysis, in many cases giving only poor resolution. However, it is clear from the energy spectral distribution figures which have been published that the energy spectra produced by different accelerators vary considerably. The narrowest spectra recorded are those by Bowen *et al.* (1946), who observed an 8% energy spread to half intensity at a peak energy of 0·6 MeV., Fry *et al.* (1947, 1948) who obtained spectra widths of 11% at 0·5 MeV. and 5·4% at 4·0 MeV. measured to half beam intensity, and Mills (1948), who obtained a spectrum width of 10% at 1·1 MeV. Much wider energy spectra than this have been reported by Hereford (1947) and Ginzton *et al.* (1948). While any desired part of the available spectrum may be selected for use by a magnetic field and suitable filters, a wide energy spectrum is clearly undesirable because of the power wasted in the unwanted part of the spectrum and the high radiation background level which it will produce.

Similarly, wide variations have been observed in the peak intensity of the accelerated beam. Little difficulty would be expected in passing quite large peak currents through very short accelerators, and this has been found to be true by most workers. The same is not necessarily true for longer machines since the effects of phase oscillations and radial defocusing forces may be of greater importance in them. Very few measurements of accelerated beam currents have been published giving quantitative information about these effects. Fry *et al.* (1947) have shown that for a travelling wave type accelerator designed to work in the bunching region electrons may be collected over almost the complete cycle. The beam current passing through 40 cm. of accelerator was measured in the absence of the R.F. field and after acceleration with the R.F. field switched on. The beam currents were found to be 2·1 and 1·8 μa. respectively. They

also found (1948) that for electrons injected at 45 kv. and an axial electric field of 25 kv/cm. an axial magnetic field of 540 gauss max. in accordance with the formula given in § 2.2 overcame the radial defocusing forces. In an accelerator two metres long, in which the bunched electrons were brought from 45° ahead of the peak field to 35° in the first 40 cm. and maintained there for the remainder of the length, the difference in accelerated peak currents was more than ten times greater with the magnetic field on than with it off. With the magnetic field on it was possible to extract 30% of the available R.F. power from the accelerating field into the electron beam, as measured by a water calorimeter. This corresponded to a peak current of 150 ma. at 4 MeV., the diameter of the energizing electron beam being approximately 7 mm. If large accelerated currents are required it is important to have an axial magnetic field for the time the electron velocity is appreciably less than that of light to counteract the radial component of the electric field. This does not seem to have been recognized by some workers.

For the accelerators about which sufficient information has been published, their efficiencies, expressed in megohms/metre, have been tabulated in column (11) of Table 3. Efficiencies measured in this way show how well the available power has been used to establish the axial R.F. field. It has been estimated in §§ 3.2 and 3.3 that for any form of accelerator there are maximum usable section lengths. When lengths short compared with this maximum are used the efficiencies will be low, improving as the length is increased. This is well illustrated by the three results of Ginzton *et al.*, in which the accelerator length has been increased in three stages and the efficiency has increased from 4·2 to 14 megohms/metre.

For the standing wave accelerators the highest effective shunt impedance has been obtained by Mills (1948) for the single resonant section and by Allen and Symonds (1947) for a three-cavity section. The latter achieved this by using cavities approximately $\lambda/2$ long loaded with drift tubes to reduce the physical length of the accelerating gaps. This reduces the transit time of the particles across the gaps, giving improved efficiency compared with that obtained with any $\lambda/2$ cavity which is not re-entrant.

However, very few workers have produced accelerators with effective shunt impedances as high as those given in Tables 1 and 2. The reason for this is that it has been found convenient to try out the main principles of acceleration on less efficient sections and so avoid the frequency stabilization and mode separation problems which only assume importance as the optimum length is approached. An increase in accelerator efficiencies is therefore to be expected in the future.

§ 9. FUTURE DEVELOPMENT OF ELECTRON ACCELERATOR

In addition to the improved efficiencies which may be expected from the use of waveguide systems designed to work nearer their optimum conditions as outlined in §§ 3.2 and 3.3, it is to be anticipated that new methods of reducing the power wastage which now occurs in any microwave accelerator will be proposed. Two methods of doing this for a travelling wave system have already been suggested by Harvie (1948b, 1949).

He suggests feeding the power passing out of the end of a travelling wave section back to the input of the section through a phase-adjusting unit to a bridge circuit designed to prevent passage into the R.F. source. This bridge circuit

Table 3. Reported Performance Data of Linear Electron Accelerators

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Bowen, Pulley and Gooden (1946)	Standing wave π-mode. One re-entrant cavity.	0·6	50	50	1000	1200	0	10	—	—
Allen and Symonds (1947)	Standing wave π-mode. One section of 3 re-entrant cavities. Mode separation — %.	0·85	—	8	2800	3000	25	12·4	0·5	12·0
Hereford (1947).	Standing wave π-mode. One section of one re-entrant cavity.									
	(*a*) One acc. through gap.	(*a*) 0·65 ± 10%	(*a*) 300–750*	70	4100	400	0	30	0·5	2·8
	(*b*) Two acc. through same gap.	(*b*) 1·0 (?)	(*b*) 300–1,250*	—	4100	400	0	30	0·5	
Fry, Harvie, Mullett and Walkinshaw (1947).	Travelling wave. One section.	0·53	65	36	10000	3000	45	40·0	2·0	0·35
Cullen and Greig (1948).	Standing wave π-mode. One section of 3 cavities. Mode separation — %.	0·3	—	—	4900	2800	2	—	0·8	—
Ginzton, Hansen and Kennedy (1948).	Travelling wave. One section.	1·7	(i) ‡	—		2860	78	(i) 91	0·75	4·2
		4·5	(ii) ‡	—		2860	78	(ii) 275	0·75	10·0
		6–6·5 †	(iii) ‡	—				(iii) 365	0·75	14·0
Lawton and Hahn (1948).	Standing wave $\pi/2$ mode. One section of									
	(*a*) 2 cavities. Mode separation 2·3%.	0·39	—	—		2800	58	8	0·75 (?)	2·5
	(*b*) 7 cavities. Mode separation 0·7%.	0·57	—	—		—	57	32	0·75 (?)	1·35
Schultz *et al.* (1948).	Standing wave π-mode. Two sections each of one cavity.	1·5 (?)	—	—	—	587	—	30	—	—
Fry, Harvie, Mullett and Walkinshaw (1948).	Travelling wave. One section.	4·0	220	150	2500	3000	1600	200	2·0	4·0
Newbery and Willshaw (1948).	Standing wave π-mode. One section of 24 cavities. Mode separation 0·2%.	—	—	—	1000	3000	—	120	0·6	—
Slater (1948 b).	Standing wave π-mode. Five sections each of 24 cavities. Mode separation 0·1%.	(Est.) 28	—	—	—	3000	2000	640	16	7·7
Snoddy and Beams (1948).	Standing wave π-mode. One section of one cavity.	1·2	—	—	—	400	—	—	—	—
Mills (1948).	Standing wave π-mode. One section of one cavity.	1·1	220 approx.	70	1000	1200	10	10	0·5	24

* To zero intensity. † Measured at maximum energy. ‡ Not known accurately.

(1) Workers. (2) Type of accelerator. (3) Measured energy at maximum intensity (MEV.). (4) Energy spectrum width between half intensity points (kev.). (5) Mean current in the pulse (ma.). (6) Duty cycle. (7) Radio frequency (Mc/s.). (8) Injection voltage (kv.). (9) Approx. length of R.F. systems (cm.). (10) Mean R.F. pulse from source (MW.). (11) Efficiency $\bar{V}^2/PL$ (MΩ/m.).

may be designed for any given ratio of power entering to power leaving the accelerator within quite a wide range, and may therefore be made to suit an accelerator with a known attenuation. If this is done it should be practicable to make accelerators with large effective shunt impedances free from the attendant troubles of frequency stability discussed in § 3.2.

The other method of improving efficiency proposed is the use of an anisotropic dielectric to load the waveguide instead of the metal irises described above. This dielectric is to have a higher effective dielectric constant in the radial than in the axial direction and may be made up of a number of spaced ceramic discs. This method of making an accelerator may have several advantages compared with one using a metal corrugated waveguide. The physical diameter of the accelerator would be smaller, and the manufacture, using ground ceramic discs, would be cheaper to make than one requiring the complicated metal to metal contacts of the corrugated guide. Small adjustments to the phase velocity should also be practicable by moving the ceramic discs. In addition to these practical advantages it should be possible to design accelerators to have higher efficiencies since the reduction in copper losses more than outweighs the increased dielectric losses.

PART II—HEAVY PARTICLE ACCELERATORS

§ 10. A STANDING-WAVE PROTON ACCELERATOR

Compared with the recent effort which has gone into the study of electron acceleration by linear methods, the field of proton acceleration is relatively unexplored. The Berkeley accelerator (Alvarez *et al.* 1947) is the only project which has so far reported the successful production of high energy protons using modern waveguide techniques. There is a consequent lack of detailed published information in this field.

Fundamentally the basic principles for the acceleration of slow particles is the same as that described earlier in this Report for electron acceleration, but there now appear some new factors which make the problem more difficult. Firstly, the type of waveguide loading which was previously found suitable for electron acceleration at relativistic velocities can be shown to be inefficient and unpracticable for slow particles. This is clearly demonstrated by a consideration of the system of collinear resonators with infinitely small coupling holes described in § 3.1. The last factor in the expression for the shunt impedance must now be replaced by $[(\sin \pi/n\beta)/(\pi/n\beta)]^2$, in order to take into account the longer transit time of the particle in traversing each resonator. The optimum number of cavities per free space wavelength is approximately $3/n\beta$. Using this number, the shunt impedance is therefore, from (11), directly proportional to the particle velocity. For protons at energies less than 20 MEV. a structure of this sort will be between 5 and 100 times less efficient than in the relativistic range.

In the Berkeley proton accelerator high ohmic loss is avoided by using a single long resonator excited in the lowest mode with an axial electric field and circular magnetic field. The transit time of the particle along the length of the resonator is long compared to the period of the R.F., and it is therefore necessary to provide drift tubes to shield the particle from the field when it is in the wrong phase. The resonator is shown in Figure 10, and can be regarded as a system of collinear re-entrant cavities all in the same phase, with the intervening walls (indicated by a broken line in the figure) removed. The power loss per unit

length of this system is then almost independent of the number of accelerating gaps. On the other hand the voltage gained by the particle is proportional to the number of gaps, and hence the shunt impedance varies inversely as the square of the particle velocity.

R.-S.-Harvie (1948) has drawn attention to this fundamental difference between behaviours of those structures suitable for slow particles and of those for relativistic particles. He points out that the Sloan and Lawrence accelerator for mercury ions has an efficiency far in excess of modern electron accelerators, which can be attributed to the fact that only one resonant circuit is used, but its voltage is applied a large number of times to the particle during its transit.

The second point of fundamental importance in slow particle acceleration is the choice of frequency. In § 3.1 it has been shown that the shunt impedance varies inversely with the square root of the wavelength, provided the hole for the passage of the particles is negligibly small compared with the wavelength. This statement must now be modified to take into account the finite hole size, which has a greater influence on efficiency when the particles are slow. Consider

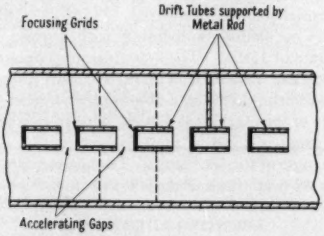


Figure 10. Schematic diagram of Berkeley proton accelerator.

an infinitely narrow voltage gap in a cylindrical tube of small dimension. At the surface of the cylinder the field is concentrated at the gap, but on the tube axis it spreads out and is attenuated to 10% of its peak value at a distance of 1·4 times the tube radius from the centre of the gap (Wang 1945). If the particle of velocity v is to traverse this distance in less than half the period of the R.F. then the radius of the cylinder must be less than $\beta\lambda/5{\cdot}6$, where λ is the free space wavelength and β is the ratio of particle velocity to that of light. For protons initially at an energy of 1 MEV. and for a 2 cm. diameter tube Cowhig (1946) estimates in this way that the wavelength must be at least one metre. For heavier ions this transit time effect will necessitate even longer operating wavelengths.

In an actual accelerator the particle will pass a sequence of voltage gaps, and their total effect has to be integrated. If this is carried out, we arrive at an equally instructive way of viewing the transit time effect. As described earlier, analysis of the field in the tube reveals a component wave with a phase velocity equal to the particle velocity. This field is given analytically by equation (7). For slow phase velocities the axial field has a minimum at the axis and increases with radial distance. Thus, as above, the voltage gaps are more efficient at the tube surface, where the field spread is not so serious.

Injection for the Berkeley proton accelerator is at 4 MeV. from a Van de Graaff generator (Turner *et al.* 1948). The drift tube diameters are fixed by the finite spread of the proton beam and are required to be at least 2 in. in diameter. The resonant frequency has therefore been chosen at 200 Mc/s. The total length of the resonator (Bradner *et al.* 1948) is 40 feet and its diameter is 38 in., with a shunt impedance of 330 megohms and a Q of 70,000. The outer diameters of the drift tubes are chosen so that the resonant frequency of the cavity remains constant along its length. The peak power supplied is 1·5 MW. from 23 triode oscillators.

The loaded guide will have a spectrum of resonant frequencies within its finite pass band, and in this case it is operated at its low frequency cut-off (zero-mode). Mode separation has, however, presented no serious trouble (Panofsky *et al.* 1948), but the problem of obtaining a uniform field distribution along the resonator by proper choice of drift tube dimensions has been more difficult. No problem is encountered in phasing the oscillators (Baker *et al.* 1948), which automatically lock into the resonant frequency of the cavity. Serber (1948) has shown analytically that with tungsten grids over the entrances to each drift tube both axial and radial stability is ensured.

The main bank of oscillators is operated with a pulse length of 300 μsec. at a pulse repetition rate of 15 c/s. The system is pre-excited by three oscillators at the output end of the resonator with a 450 μsec. pulse. The remaining oscillators are coupled to the cavity by a 50-ohm line, one wavelength long.

The output energy of the machine is 32 MeV. with an energy spread of $\pm$ 100 kv. The mean output current is 2×10^{-10} amp., which, with a duty cycle of 300 : 1, gives a peak pulse current of 6×10^{-8} amp. The pulse current injected from the Van de Graaff generator is approximately 100 μa.

ACKNOWLEDGMENTS

The authors wish to thank the Director of the Atomic Energy Research Establishment, Harwell, for permission to publish this paper. They are also indebted to Mr. R. B. R.-S.-Harvie and Mr. F. K. Goward for much helpful discussion and criticism.

GENERAL REFERENCES

Blewitt, J. P., 1946, *Phys. Rev.*, **69**, 87.
Brobeck, W. H., Lawrence, E. O., McKenzie, K. R., McMillan, E. M., Serber, R., Sewell, D. C., Simpson, K. M., and Thornton, R. L., 1947, *Phys. Rev.*, **71**, 449.
Cockcroft, J. D., and Walton, E. T. S., 1930, *Proc. Roy. Soc.* A, **129**, 477; 1932, *Ibid.*, **136**, 619.
Kerst, D. W., 1941, *Phys. Rev.*, **60**, 47.
Lauritsen, C. C., and Bennett, R. D., 1928, *Phys. Rev.*, **32**, 850.
Lawrence, E. O., and Edlefson, N. E., 1930, *Science*, **72**, 376.
Lawrence, E. O., and Livingston, M. S., 1931, *Phys. Rev.*, **37**, 1707.
McMillan, E. M., 1945, *Phys. Rev.*, **68**, 143.
Morrison, P. J., 1947, *Appl. Phys.*, **18**, 133.
Oliphant, M. L., Gooden, J. S., and Hide, G. S., 1947, *Proc. Phys. Soc.*, **49**, 666.
Page, L., and Adams, N. I., 1940, *Electrodynamics* (D. Van Nostrand Co., Inc.).
Post, R. F., 1946, *Phys. Rev.*, **69**, 126.
Schiff, L. I., 1946, *Rev. Sci. Instrum.*, **17**, 6.
Van de Graaff, R. J., 1931, *Phys. Rev.*, **38**, 1919.
Van de Graaff, R. J., Trump, J. G., and Buechner, W. W., 1946–47, *Rep. Prog. Phys.*, Vol. XI (London : Physical Society), p. 1.

VEKSLER, V. K., 1945, *J. Phys. Acad. Sci. U.S.S.R.*, **9**, 153.
WANG, C. C., 1945, *J. Appl. Phys.*, **16**, 351.
WESTENDORP, W. F., and CHARLTON, E. E., 1945, *J. Appl. Phys.*, **16**, 581.
WOODYARD, J. R., 1948, *Elect. Engng.*, **67**, 759.

LINEAR ELECTRON ACCELERATOR REFERENCES

AKELEY, E. S., 1946 a, *J. Appl. Phys.*, **17**, 1056; 1946 b, *Phys. Rev.*, **69**, 50; 1946 c, *Ibid.*, **69**, 255.
ALLEN, W. D., and SYMONDS, J. L., 1947, *Proc. Phys. Soc.*, **59**, 622.
BEAMS, J. W., and SNODDY, L. B., 1933, *Phys. Rev.*, **44**, 784.
BEAMS, J. W., and TROTTER, H., Jr., 1934, *Phys. Rev.*, **45**, 849.
BOWEN, E. G., PULLEY, O. O., and GOODEN, J. S., 1946, *Nature, Lond.*, **157**, 840.
BRILLOUIN, L., 1947, *Phys. Rev.*, **71**, 483; 1948, *Ibid.*, **74**, 90.
BRUCK, G. G., and WICKER, E. R., 1947, *J. Appl. Phys.*, **18**, 766.
CHU, E. L., and HANSEN, W. W., 1947, *J. Appl. Phys.*, **18**, 996.
COWHIG, W. T., 1947, unpublished work.
COWHIG, W. T., and JONES, E. J., 1947, *T.R.E. Report* A.E. 1008.
CULLEN, A. B., and GREIG, J. A., 1948, *J. Appl. Phys.*, **19**, 47.
CUTLER, C. C., 1944, *Bell Telephone Labs. Report* MM-44-160-218.
DEVONS, S., 1947, private communication.
FRANKEL, S., 1947, *J. Appl. Phys.*, **18**, 650.
FRY, D. W., R.-S.-HARVIE, R. B., MULLETT, L. B., and WALKINSHAW, W., 1947, *Nature, Lond.*, **160**, 351; 1948, *Ibid.*, **162**, 859.
GABOR, D., 1947 a, *Nature, Lond.*, **159**, 303.
GINZTON, E. L., HANSEN, W. W., and KENNEDY, W. R., 1948, *Rev. Sci. Instrum.*, **19**, 89.
HALPERN, J., EVERHART, E., RAPUANO, R. A., and SLATER, J. C., 1946, *Phys. Rev.*, **69**, 688.
HANSEN, W. W., 1948, Paper read to Conference at M.I.T., Mass., June.
HANSEN, W. W., and RICHTMYER, R. D., 1939, *J. Appl. Phys.*, **10**, 189.
HAXBY, R. O., AKELEY, E. S., GINZBERG, A., SMITH, R. N., WELCH, H. W., and WHALEY, R. M., 1946, *Phys. Rev.*, **70**, 797.
HEREFORD, F. L., 1947, *J. Appl. Phys.*, **18**, 956.
R.-S.-HARVIE, R. B., 1945, *T.R.E. Memorandum* S3/M4601; 1948 a, *Proc. Phys. Soc.*, **61**, 255; 1948 b, *Nature, Lond.*, **162**, 890; 1949, *Proc. Phys. Soc.* B, **62**, 270.
HUDSPETH, E. L., 1946, *Phys. Rev.*, **69**, 671.
LAWTON, E. J., 1948, *J. Appl. Phys.*, **19**, 534.
LAWTON, E. J., and HAHN, W. C., 1948, *J. Appl. Phys.*, **19**, 642.
MILLS, B. Y., 1948, *Report No.* 77 of Radio Physics Laboratory, University of Sydney, Australia.
MULLETT, L. B., 1946 a, *T.R.E. Report* T.2028; 1946 b, *Ibid.*, T. 2021.
MULLETT, L. B., and LOACH, B., 1948, *Proc. Phys. Soc.*, **61**, 271.
NEWBERY, G. R., 1948 a, *G.E.C. Research Laboratories Memorandum* dated 23rd January; 1948 b, *The Microwave Linear Electron Accelerator* (*Brit. J. Radiology*), in press.
NEWBERY, G. R., and WILLSHAW, W. E., 1948, *Nature, Lond.*, **161**, 519.
OLINER, A. A., 1948, *J. Appl. Phys.*, **19**, 109.
SCHULTZ, H. L., 1948, Paper read to Conference at M.I.T., Massachusetts, June.
SCHULTZ, H. L., BERINGER, R., CLARKE, C. L., LOCKWOOD, J. A., MCCARTHY, R. L., MONTGOMERY, C. G., RICE, P. J., and WATSON, W. W., 1947, *Phys. Rev.*, **72**, 346.
SLATER, J. C., 1946, *Phys. Rev.*, **70**, 799; 1948 a, *Rev. Mod. Phys.*, **20**, 473; 1948 b, private communication.
SLATER, J. C., and FRANK, N. H., 1933, *Introduction to Theoretical Physics* (New York: McGraw-Hill).
SNODDY, L. B., and BEAMS, J. W., 1948, *Phys. Rev.*, **74**, 126 A.
SNODDY, L. B., TROTTER, H., HAM, W., and BEAMS, J. W., 1937, *J. Franklin Inst.*, **223**, 55.
STARR, A. T., 1945, *T.R.E. Memorandum* dated 22nd November.
STRATTON, J. A., 1941, *Electromagnetic Theory* (New York: McGraw-Hill).
WALKINSHAW, W., 1947, *T.R.E. Report* A.E.1005; 1948, *Proc. Phys. Soc.*, **61**, 246.

WILLSHAW, W. E., 1946, *G.E.C. Research Laboratories, England. Memorandum* dated 31st December; 1947 a, *Ibid.*, dated 16th January; 1947 b, *Ibid.*, dated 3rd March; 1948, *Ibid.*, dated 1st January.

WILLSHAW, W. E., and LAMONT, H. R. L., 1945, *Confidential Report from Research Labs. of G.E.C. Ltd., England, No. 8766 B.*

WOODYARD, J. R., 1946, *Phys. Rev.*, **69**, 50.

HEAVY PARTICLE LINEAR ACCELERATOR REFERENCES

ALVAREZ, L. W., 1946, *Phys. Rev.*, **70**, 799.

ALVAREZ, L. W., BRADNER, H., GORDON, H., MARSHALL, L. C., OPPENHEIMER, F., PANOFSKY, W., RICHMAN, C., SERBER, R., TURNER, C., and WOODYARD, J. R., 1947, *Science*, **106**, 506.

BAKER, W. R., FRANK, J. V., and GOW, J. D., 1948, *Phys. Rev.*, **73**, 535 A.

BEAMS, J. W., and SNODDY, L. B., 1934, *Phys. Rev.*, **45**, 287.

BRADNER, H., CRAWFORD, R., GORDON, H., and WOODYARD, J. R., 1948, *Phys. Rev.*, **73**, 534 A.

COWHIG, W. T., 1946, *T.R.E. Memorandum* S1/M178.

GABOR, D., 1947 b, *Nature, Lond.*, **160**, 89.

LAWRENCE, E. O., and SLOAN, D. H., 1931, *Proc. Nat. Acad. Sci.*, **17**, 64.

OPPENHEIMER, F., JOHNSTON, L. H., and RICHMAN, C., 1946, *Phys. Rev.* **70**, 447 A.

PANOFSKY, W. K. H., 1946, *Phys. Rev.*, **70**, 447 A.

PANOFSKY, W. K. H., RICHMAN, C., and OPPENHEIMER, F., 1948, *Phys. Rev.*, **73**, 535 A.

SERBER, R., 1948, *Phys. Rev.*, **73**, 535 A.

SLOAN, D. H., and COATES, W. M., 1934, *Phys. Rev.*, **46**, 539.

SLOAN, D. H., and LAWRENCE, E. O., 1931, *Phys. Rev.*, **38**, 2021.

TURNER, C., CORK, B., BALLAM, J., and GORDON, H., 1948, *Phys. Rev.*, **73**, 534 A.

WIDEROE, R., 1928, *Archiv Electrotech.*, **21**, 387.

WOODYARD, J. R., MARTINELLI, E. A., TOULIS, W., and PANOFSKY, W. K. H., 1946, *Phys. Rev.*, **70**, 447.

VISCOSITY AND RELATED PROPERTIES IN GLASS

By G. O. JONES
The Clarendon Laboratory, Oxford

(Originally presented before the Society as a Lecture Survey on 15th February 1946)

CONTENTS

	PAGE
§ 1. Introduction	133
§ 2. Structure and general properties of glass	134
§ 3. Configurational changes in liquids and amorphous solids	135
§ 4. Transformation phenomena in glass	136
§ 5. The flow properties of glass	139
§ 6. Viscosity	141
§ 7. Elasticity	145
§ 8. Interpretation of the flow data	148
(i) Representation by mechanical models	148
(ii) Significance of a distribution of relaxation times	152
(iii) Activation energy of viscous flow	154
(iv) Further evidence for configurational adjustments below the transformation region	155
(v) Other features of interest	155
§ 9. Practical aspects of the data in the transformation region	156
§ 10. Practical aspects at ordinary temperatures	158
References	161

§ 1. INTRODUCTION

THE flow properties of glass are of special interest for a number of reasons. They are of theoretical interest because of the importance of flow properties in the development of the theory of liquids, and because glass can be regarded as a liquid even at temperatures where it fits more closely the familiar conception of a solid. It is in fact an especially informing type of liquid, because particular aspects of its behaviour can, as it were, be brought into *focus* by suitably choosing the temperature.

These properties are of obvious practical importance. The special qualities of glass make it virtually the only possible material for many practical applications, and the demands then made from the purely mechanical or structural points of view are often severe. It is a rather remarkable fact that glass, a thermodynamically unstable liquid at room temperature, is actually used where the greatest rigidity and permanence are required—in the construction of large astronomical mirrors.

In the present report a survey is given of existing information on these flow properties, with particular reference to the behaviour in and below the transformation temperature, and a number of the theoretical and practical implications of the data are discussed. A short account is first given of the structure and general properties of glass, and in particular of the significance of the transformation phenomena.

§2. STRUCTURE AND GENERAL PROPERTIES OF GLASS

The following definition given by Morey is a convenient starting-point: "A glass is an inorganic* substance in a condition which is continuous with, and analogous to, the liquid state of that substance, but which, as the result of having been cooled from a fused condition, has attained so high a degree of viscosity as to be for all practical purposes rigid". Thus although glass has many properties characteristic of ordinary solids, such as rigidity, hardness, mechanical strength, etc., it nevertheless exhibits true liquid properties such as viscous flow also.

The x-ray diffraction patterns consist of a small number of diffuse rings, indicating that the immediate surroundings of any atom are nearly similar to those in the corresponding crystalline structure (where such a structure exists), but the more distant arrangement becomes irregular and varies from one atom to another. (In vitreous silica, for example, tetrahedral SiO_4 groups exist, linked together more or less at random.) The structure thus has a degree of short-range order, but long-range disorder; broadly speaking it may be likened to that which would be shown by an "instantaneous" picture of the structure of an ordinary liquid. Atomic migrations occur, however, relatively much less frequently, so that processes such as diffusion or viscous flow are much slower.

We shall not discuss in any detail the reasons why particular substances have a tendency to supercool and form glasses. Many organic substances may be said to form glasses on cooling because no crystalline structure can form on account of the length or complicated nature of the molecules. The vitreous state is much rarer among inorganic substances; the conditions under which relatively stable irregular three-dimensional frameworks can be formed have been discussed by Warren (1933), Zachariasen (1932) and others. They can occur, for example, where metallic or homopolar structures have approximately the same energy, as in the case of the element selenium. In the inorganic glasses based on silica the strongest bonds are those between silicon and oxygen, and the main factor in their glass-forming ability appears to be that the angles between the two Si–O bonds of one oxygen atom may vary within quite wide limits without loss of stability. Here the bond is roughly half ionic and half homopolar, that is, the bond energy is again about equally divided between a bond of a non-directed kind and an essentially directed bond. A simple statement of the main obvious condition for the formation of a glass is that the viscosity should be high at the crystallization temperature, so that crystallization can take place only slowly. In the case of silica, for instance, the viscosity is so high at the melting point that crystals of silica can hardly be seen to melt, but appear only to sinter slowly. For similar reasons very many crystalline modifications of silica exist, and many of the possible transitions are "sluggish".

Below the liquidus temperature corresponding to the particular composition of any glass the structure is of course thermodynamically unstable, and a tendency towards crystallization (or "devitrification") exists. As pointed out by Tammann (1933) and Becker (1938), the rate of crystallization is determined by the probability of formation of stable crystal nuclei as well as by the various factors influencing their rate of growth. These all depend upon temperature, and there is in fact a dangerous range of temperature for crystallization. Provided the glass has been cooled fairly rapidly through this range it will remain fairly stable,

* The exclusion of organic glasses from this definition is no more than conventional; in the present paper, however, the discussion is mainly limited to inorganic glasses.

with only a very slow rate of devitrification. In this region glass is a supercooled liquid; crystallization is a localized process starting from nuclei, and non-crystalline portions exhibit properties which are continuous with those above the liquidus temperature.

To complete this very brief account of the general properties of glass, we mention the influence of changing composition on the properties of the liquid. From the physico-chemical point of view ordinary glass is a supercooled solution of molten oxides. The effect of "fluxes" such as soda or lime is of course to lower the liquidus temperature, so that homogeneous glasses may be manufactured without the need for excessively high melting temperatures. These substances also lower the viscosity of the glass at a particular temperature; this is not due simply to the lower strength of, say, the Na–O bond, but to the reduced co-ordination of the structure containing a proportion of sodium or calcium atoms substituted for silicon.

In plotting the variation of some property against composition it is frequently found that a maximum or minimum occurs at the composition which corresponds to that of a known compound, say some particular silicate. This not unexpected result of the greater probability of atomic arrangements of lower free energy has often been exaggerated in importance and referred to as evidence for the existence of "compounds" in glass.

§3. CONFIGURATIONAL CHANGES IN LIQUIDS AND AMORPHOUS SOLIDS

Modern theories of the liquid state (see for instance Bernal 1937, Frenkel 1946, Mott and Gurney 1939) incorporate the idea that the degree of configurational order in a liquid is a function of the temperature, decreasing with increasing temperature. Thus when the temperature is increased the configuration must be changed to one of greater disorder, and a part of the heat supplied must be employed for this purpose. It follows from the relationship which exists between entropy and disorder that a positive quantity is required.

There is thus a configurational contribution to the specific heat of liquids in addition to the contribution required to increase the vibrational energy, and this is probably the main reason for the increased specific heat of a liquid near its melting point as compared with that of the solid.

These ideas, which have been developed especially by Bernal, were first introduced by Simon (1931) to explain the sudden increase (with increasing temperature) observed in the specific heat of supercooled glycerol at a temperature of about $-100°$ C. He suggested that below this temperature the equilibrium times of configurational changes became so long that they would not occur during ordinary measurements, and the measured value of the specific heat would accordingly be smaller by the corresponding amount.

More or less sudden changes are in fact shown in the physical properties of all supercooled liquids or amorphous solids at a particular temperature—about $-100°$ for glycerol, $-70°$ for rubber, 200° for boric oxide, and 500° for ordinary inorganic glasses. It was shown by Samsoën and by Tammann (1933) that the changes always occurred at temperatures at which the viscosity was of the order of 10^{13} poises. This has an important significance, since this value of the viscosity corresponds to a relaxation time for the release of ordinary mechanical stresses of seconds or minutes, that is, of the sort of time taken in ordinary measurements.

(The relaxation time is defined, after Maxwell, as the time in which a stress falls to $1/e$ of its original value if the strain is kept constant; it is equal to η/E, where η is the coefficient of viscosity and E the modulus of rigidity.) The equilibrium time of configurational changes might be expected to be of the same order.

The temperature range over which these changes in properties are observed experimentally is known as the "transformation region" (or sometimes "transition region"). The phenomena of the transformation region are extremely important in glass. The equilibrium times of configurational changes are in fact so long that in ordinary glasses at ordinary temperatures we always have a material possessing a configuration which really corresponds to a much higher temperature, but vibrational energy roughly appropriate to its actual temperature. For this reason the transformation phenomena in glass must be described here in some detail.

§4. TRANSFORMATION PHENOMENA IN GLASS

The phenomena of the transformation region may be most conveniently described by considering the variation with temperature of some physical property (or log (property)) such as the specific volume, total heat, or log (viscosity), as would be observed in experiments carried out with the temperature falling at uniform rate. The behaviour is illustrated diagrammatically in Figure 1. It will

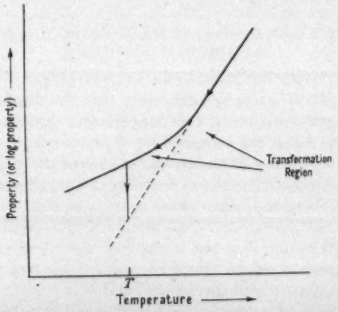


Figure 1. Illustration of typical behaviour in the transformation region.

be seen that the slope changes more or less suddenly at the transformation region since below this region there is no further alteration in the configurational contribution to the value of the property. If, however, the fall in temperature is halted at T, say, and the temperature maintained constant, the value of the property slowly moves to an equilibrium value represented by a point on the dotted line, and there is an evolution of heat corresponding to the decrease in internal energy. The dotted line is the extrapolation of the high temperature part of the curve, so that the glass has now arrived again at the ordinary supercooled liquid or "metastable" condition. If the experiment is carried out sufficiently slowly, the dotted line is of course followed throughout. This approach to the equilibrium condition at any temperature is generally known as "stabilization"; it takes longer and longer times as the temperature is lowered, glass in ordinary use being always incompletely stabilized, as already indicated.

It will be clear that the exact location of the transformation region as experimentally determined in this way must depend on the rate of cooling, higher rates of cooling causing the transformation phenomena to be shown at higher temperatures. Also, the measured value of any physical property in this region will depend on the rate of heating or cooling during the measurements and on the previous thermal history in the same temperature region. Measurements carried out at much lower temperatures will depend only on the actual temperature and on the previous thermal history in the transformation region, particularly the rate of cooling here, since this determines the configuration. It follows from these considerations also that measurements of property–temperature relationships made with a rising temperature schedule between ordinary temperatures and the transformation region immediately following measurements obtained with a falling temperature schedule will give approximately similar results.

If derivative properties such as the specific heat or the coefficient of thermal expansion are plotted against temperature, the curve of course shows not simply a change in slope, but a more or less abrupt discontinuity; we have already mentioned the sudden increase in the specific heat of glycerol and its explanation by Simon (1931).

When the temperature of a specimen of glass is varied above the transformation region, the change in the configurational contribution to any physical property will nearly always have the same sign as the change in the thermal energy contribution. (There are notable exceptions to this rule in the properties of silica and water.) The condition of glass cooled through its transformation region is that it appears to possess properties "frozen-in" from a temperature higher than its actual temperature; thus it usually has a higher specific volume and a lower viscosity than it would have in equilibrium at its actual temperature. As an illustration we may add that it is easily possible by adjusting the rate of cooling to produce specimens of glass which differ in these properties by amounts corresponding to as much as 100° in terms of "configurational temperature". It is also possible in the transformation region for glass to exhibit properties which appear to be "frozen-in" from a slightly lower temperature, since the configuration will not immediately follow a rapid increase of temperature in this region. The final equilibrium condition is however always the same at a particular temperature from whatever side it is approached.

For convenience in practice a "transformation temperature", T_g, is defined as the temperature at which a fairly sharp increase in slope occurs in the measurement of the thermal expansion of a rod of the particular glass heated with a rising temperature schedule of 4° per minute. This approximately coincides with the transformation temperature as defined in any other way based on the observation of sudden changes in physical properties or their derivatives with changing temperature, the viscosity being between 10^{13} and 10^{14} poises. The conventional definition is useful for many purposes, though without theoretical significance. The practical importance of T_g is that this temperature is at about the middle of the fairly narrow range of temperature into which the glass is heated for the purpose of annealing, the relaxation time for the release of stresses being of the order of seconds or minutes, as already stated, or conveniently short. For convenience, a "softening temperature", M_g, is also defined from the results of expansion measurements as, in effect, the temperature above which it is impossible to carry out ordinary measurements by conventional methods because of sag or

yield in the specimen; the viscosity here is about 10^{11}–10^{12} poises (relaxation time about 1 second). This may be regarded as the extreme upper limit of the annealing and the transformation regions, which are more or less coincident. Annealing schedules for a particular glass can usually be prepared by studying the results of thermal expansion measurements.

The slow approach to the equilibrium condition below the transformation region has been observed in measurements of many of the physical properties of glass, such as the viscosity (Stott and co-workers 1925, Lillie 1933), refractive index (Winter 1943 and others), specific volume (Morey and Merwin 1932), electrical conductivity (Littleton and Wetmore 1936), specific heat (of glycerol, Oblad and Newton 1936), etc. The approach to a higher temperature configuration after a sudden raising in temperature in the annealing region has also been observed by Lillie, for the property of viscosity, in an important series of experiments to be discussed later in this paper, by Tool and co-workers (1920) and others for certain thermal properties, and by Winter (1943) for the refractive index.

These processes have not yet however received adequate quantitative description or explanation as there have been many complicating factors in the experiments so far carried out. A start has been made by Tool (1942, 1946) in an empirical way; he relates the actual properties of any specimen of glass below the transformation region to its actual temperature and to its "fictive temperature", the temperature at which the actual configuration of the glass would be in equilibrium—or, as we should say, its configurational temperature. He then makes the somewhat arbitrary assumption that the fictive temperature should approach the actual temperature at a rate proportional to the reciprocal of the viscosity and the difference between the two temperatures; he does not, however, appear to realize clearly the real significance of the fictive temperature, and his approximate treatment does not take account of the fact that the two "temperatures" may contribute in different ways to the value of a physical property.

With a proper understanding of the laws governing the transformation phenomena it would be possible to forecast the form of the curve obtained in a measurement of, say, thermal expansion, given the rate of heating during the measurements and details of the previous cooling schedule through the transformation region. At present this is not possible. Broadly speaking, of course, the reason for the increased rate of expansion above T_g is that appropriate configurational changes are now also taking place. The general form of the curve is mainly determined by the particular configuration "frozen-in" during the previous cooling. Thus a very rapidly cooled specimen is already "blown-up" to a too great volume and will expand correspondingly less when heated from a lower temperature to T_k.

Recent suggestions have been made by Winter (1943) which are not in harmony with the above description. She considers that glass is stable both below a lower limit of the transformation range or above an upper limit. It is true that the rate at which the properties of unstabilized glass change, at a temperature well below T_g, may be made very slow by a suitable annealing treatment (that is, by lowering the configurational temperature as far as possible). The present author maintains the view, however, that there is no real lower limit to the transformation region when experiments are extended over very long periods, and the transformation phenomena are all related to departures from thermodynamic

equilibrium resulting from changes of temperature which are too rapid for the structure to follow. In other words, there is no discontinuity anywhere which cannot be "smoothed out" by performing experiments sufficiently slowly.

Many workers on glass have not properly understood the significance of the transformation phenomena as explained by Simon. Morey (1938), for instance, considers that the process of stabilization consists "only in diminishing the interatomic distances, and not in the major atomic movements necessary to form new atomic arrangements". The fact that the configuration has to be changed clearly contradicts these ideas, and time-consuming migratory processes analogous to those which occur in diffusion or viscous flow must actually occur during stabilization. Much more experimental investigation is needed into the slow changes in structure and the absorptions or evolutions of heat which accompany them; for various reasons these experiments are difficult, and at lower temperatures time-consuming. It would however be of very great interest to make measurements on the physical properties of very old specimens of glass, of which many exist dating from 1500 B.C. or earlier, and of fairly standard composition, and then to use the same glass for a series of further experiments after heating. In this way one would, so to speak, at once obtain a point on the curve of property–time at $t=3{,}500$ years, though uncertainty as to the original cooling schedule would limit the usefulness of the method.

We should mention that devitrification is of course still possible even below the transformation region. Thus although the supercooled liquid condition is a metastable condition of minimum free energy, this is not the lowest minimum; if a crystalline form exists it has lower free energy. There is however an important difference between the two kinds of approach to equilibrium. Devitrification is a nucleus-sensitive phenomenon; thus, for instance, homogeneous glass often shows a greater tendency to crystallize on account of the presence of crystal or near-crystalline nuclei. The process of stabilization however proceeds continuously and uniformly throughout the material. The required rearrangements can be carried out steadily by individual migrations similar to those responsible for viscous flow.

Finally, two further points of interest (Smekal 1947) are mentioned: in some organic glasses the transformation region is actually above the normal freezing point; also, more than one "transformation temperature" can sometimes be observed in these glasses, presumably due to the "freezing" of different types of configurational changes at different temperatures.

§5. THE FLOW PROPERTIES OF GLASS

The main features of the rheological behaviour of glass may be summarized from the results of many investigations in the following two equations, if we neglect the influence of simultaneously occurring stabilization. For simplicity we consider only shear stresses and strains. Also, we shall be concerned only with typical behaviour, and not with the influence of chemical composition on the magnitudes of the various properties discussed.

Under a constant applied stress, applied at time $t=0$, the development of the strain is given by the following:

$$\text{Strain} = a + bt + cf(t), \qquad \ldots\ldots(1)$$

where a, b and c are constants proportional to the stress for a given glass and temperature. a represents an instantaneous elastic strain, and bt a strictly classical

or Newtonian viscous flow at uniform rate; the magnitudes of the rigidity modulus and the coefficient of viscosity are implicit in the values of these constants. c represents a delayed elastic strain which develops slowly according to a "decay" function of time $f(t)$, of value 0 at $t=0$, and 1 at $t=\infty$. The term "inelastic" has also been used to describe this part of the strain (see Zener 1948).

On removal of the stress, the instantaneous elastic strain is at once completely recovered, and the fraction n of the total delayed elastic strain which has been allowed to develop at the time of removal of the stress is also recovered, probably according to the same decay function. The recovery–time relationship can then be represented by the following equation (recovery being measured in strain units, and t now being measured from the time at which the stress is removed):

$$\text{Recovery} = a + ncf(t). \qquad \ldots\ldots(2)$$

The part of the original flow corresponding to true viscosity is not recovered.

The variation with temperature of these constants for a particular glass and the form of the decay function for delayed elasticity are described in detail later.

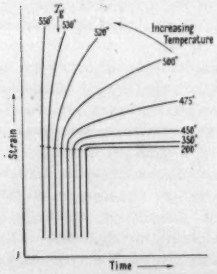


Figure 2. Comparison of typical strain-time curves under the same (constant) applied stress for a glass at different temperatures.

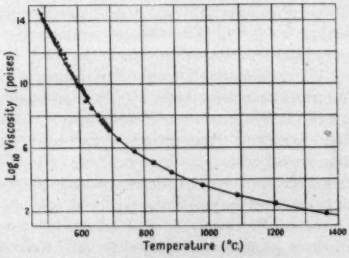


Figure 4. Viscosity–temperature relationship for a glass from the melting temperature to the transformation region. (After Lillie.)

A qualitative picture of the behaviour in and below the transformation region is given in Figure 2, where a family of curves is drawn to the same scale, representing strain–time relationships at different temperatures under the same constant applied stress for a specimen of a glass for which the position of T_g is indicated.* At temperatures near room temperature most of the observed strain consists of the instantaneous elastic strain, with a small delayed elastic strain and a very low rate of irreversible flow, if any. In the annealing or transformation region the effects of instantaneous elasticity, delayed elasticity and viscous flow are of about equal importance in the times of ordinary measurements. At temperatures well above this region the elastic strains are practically impossible to measure accurately because of the very fast rate of viscous flow. At sufficiently high temperatures only the viscous flow can be detected.

* For convenience in comparison, Figures 2, 3, 5, 7, 8, refer to specimens of the same soda–lime–silica glass for which T_g is 534°. The measurements were made by the author. Figures 4 and 6, from papers by Lillie, refer to another glass.

As the temperature is raised from room temperature above the transformation region, the structure "falls to pieces", as it were, when time is allowed, though the instantaneous value of the modulus of elasticity is practically unchanged; this illustrates very clearly the way in which liquids show real elasticity, even in shear, although the lowest value of viscosity illustrated in this figure is, of course, about 10^{14} times as high as the values shown by ordinary liquids.

The complete behaviour at a single temperature is illustrated in Figure 3, which shows the instantaneous elastic strain AB, the delayed elastic strain and viscous flow BC, and also the instantaneous recovery CD and delayed recovery

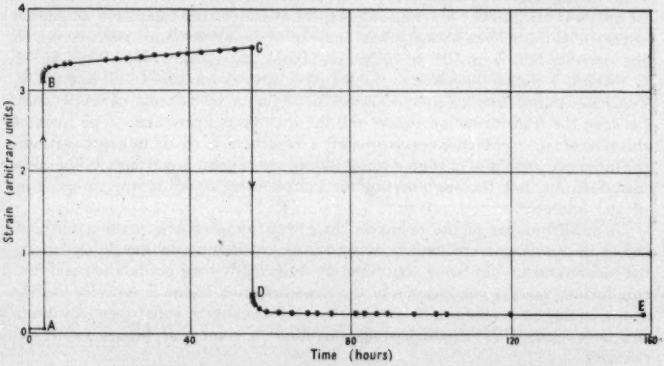


Figure 3. Strain-time behaviour under constant stress, and recovery after removal of the stress, for a glass at a temperature well below the transformation region.

DE after removal of the stress at C. In this case it was possible to continue measurements until the delayed elastic strain was practically complete, so that the limiting slope of BC represents the rate of irreversible or purely viscous flow. The form of the function $f(t)$ can be derived from the shape of the portion DE of the curve, or from the portion BC by subtracting the linear contribution due to viscous flow.

As implied in the above description, the yield point for glass is at zero stress—apart from possible negligibly small effects due to surface tension. There is no work-hardening, and the irreversible part of the flow is purely viscous, according to the strict Newtonian definition of viscosity.

§6. VISCOSITY

It is thus possible to make measurements of a true viscosity in glass down to as low temperatures as permitted by the sensitivity of the technique employed. At high temperatures, conventional methods as used for ordinary liquids (such as Margules' method, employing two concentric cylinders of which one rotates) may be used; at lower temperatures, where the value of viscosity is much higher and the glass behaves almost as a solid, the measurements are made by methods similar to those used for measuring the "creep" in metallic specimens—although even here a true viscosity is measured in the case of glass. It is however necessary at

these lower temperatures to analyse the strain–time curves carefully or else to continue measurements until the delayed elastic part of the strain is complete, as described by the author elsewhere (Jones 1944).

For wide ranges of temperature above the transformation region the viscosity varies with temperature nearly according to the function of the form $A \exp(B/\boldsymbol{RT})$ (where A and B are constants depending on the composition) derived from various points of view by Andrade (1934), Eyring (1936) and Frenkel (1946). The whole variation is in fact described by saying that the value of the activation energy B (per Avogadro number of elementary processes) for viscous flow calculated from this equation decreases slowly with increasing temperature. This seems to be true for all glasses; for boric oxide, for instance, the calculated activation energy falls from 75 to about 20 kcal/mole over the temperature range in which the viscosity falls from 10^{13} to 10^{2} poises (Parks and Spaght 1935, Ward 1937).

Figure 4 shows the variation in log (viscosity) as measured by Lillie (1930, 1931) for a particular glass by various methods over the whole range of temperature between the transformation region and the melting temperature. The form of this type of curve (which is approximately a hyperbola, as would be expected), and the influence upon it of changing composition, are of great importance to the glass manufacturer, but for our purpose the behaviour at lower temperatures is of greater interest.

A small number of the points at the lowest temperatures investigated and shown in this curve were actually derived from experiments on unstabilized glass, the measurements not being continued for sufficiently long periods to reach the equilibrium values; the location of the transformation region is not very clearly shown in this curve, but can be seen in the results of similar experiments by Stott and co-workers (1925), which were extended to somewhat higher values of viscosity.

In Figure 5* are shown the results of measurements by Jones (1944) which were continued to a still higher value of viscosity. The value obtained at 350° is that derived from the curve of Figure 3, and is probably the highest value of true viscosity yet measured directly. No attempt was made in these experiments to stabilize the glass as their object was to extend measurements to as low a temperature as possible, and it was considered inevitable that the region of instability would eventually be entered. It will be seen that there is a very marked bend in the curve; it is of interest that it occurs well below T_g, probably because of the long times normally required in measurements of viscosity at these temperatures before the delayed elasticity is sufficiently "worked out". (It should be mentioned that some authors have incorrectly described as due to the transformation phenomena other departures from linearity in property–temperature curves. Thus the curve of log (viscosity) against reciprocal temperature often appears to show a fairly sharp "bend" at a temperature appreciably higher than T_g. This is only an indication that the particular glass does not exactly obey the expected or "theoretical" Andrade-type formula, or that the activation energy calculated from this formula is not exactly constant; it has no connection with the transformation phenomena.)

* There is a slight discontinuity between the two portions of the curve taken by different methods; at the higher temperatures measurements were made of the extension of a fibre, as in nearly all other published experiments in this region, but at the lower temperatures measurements were made of the bending of centrally loaded laths of drawn sheets of the glass. The possible reasons for this discontinuity are not discussed here.

If the calculation of the activation energy for viscous flow is extended to these values of viscosity below the transformation region, the striking result is obtained that its value falls off with *decreasing* temperature; thus a maximum value is shown in the transformation region. The results of Figure 5 show a very sharp fall in the calculated activation energy from about 55 kcal/mole in the transformation region to about 6 kcal/mole at the lowest temperature investigated. These results strongly support the conclusions of Stanworth (1937), who analysed the viscosity data available in 1937, and came to the conclusion that there was evidence for a maximum value in the transformation region. As a first approximation we would expect a constant or "frozen" value of the activation energy at all temperatures below the transformation region, since this must be determined entirely by the configuration; the possible significance of the departures from this behaviour is discussed later.

It should perhaps be stressed that there is no sudden enormous increase of viscosity on passing down through the transformation region, as suggested by some

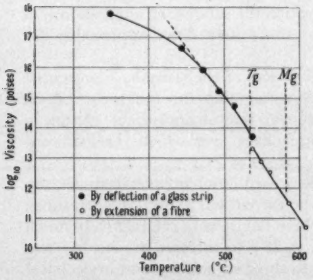


Figure 5. Variation of viscosity with temperature in and below the transformation region.

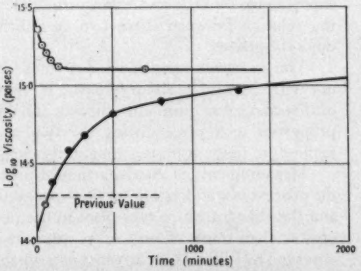


Figure 6. Viscosity–time curves for two samples of a glass at 486·7°c. The upper curve is for a sample previously heated at 477·8°c. for 64 hours; the lower curve is for a sample in the freshly drawn condition. (After Lillie.)

authors. It is the fact that glass becomes noticeably "brittle", that is, that the relaxation times become much longer than the times of ordinary operations, which has led to this common error.

The approach to the equilibrium value of the viscosity was studied by Lillie (1933), who was the first to show that the equilibrium value of a property could be approached from either side. This is illustrated in Figure 6, taken from his paper; the lower curve shows the change in viscosity with time for a freshly drawn fibre at 487°, during continuous measurement over a long period. It will be seen that the viscosity rose from 10^{14} to just over 10^{15} poises, and gave every indication of approaching an equilibrium value. The upper curve shows the change in viscosity with time for a sample at 487° which had previously been heated for a long period (64 hours) at a temperature 9° lower; it will be seen that in this case the value decreased to the same final equilibrium value. It is clear that the specimen of the upper curve had already reached a state of equilibrium corresponding to a temperature below 487°, and when raised in temperature it did not

immediately reach its new condition but did so only slowly during the measurements. However, presumably because this change corresponded only to a difference of temperature of 9° or less, it took place quite rapidly, whilst the freshly drawn, rapidly cooled sample of the lower curve, with frozen-in properties corresponding to an appreciably higher temperature, took a longer time to reach the condition corresponding to 487°. The dotted line represents the value of the viscosity which had previously been given by Lillie for the same glass, and the same temperature, and drawn as the highest value of viscosity of Figure 4, taken from an earlier paper.

The time taken for the difference of log (viscosity) to fall to $1/e$ of its original value was about 400 minutes for the case of the rapidly cooled fibre. The corresponding relaxation time for the relaxation of mechanical stresses through viscous flow would be about 10 minutes at the original value of the viscosity, or about 100 minutes at its final equilibrium value; these times thus appear to be of about the same order at temperatures for which data are available, the equilibrium time of the stabilization process being, however, probably the longer. Recent experiments by Dale and Stanworth (1945) lead to the interesting suggestion that the relation between these two equilibrium times may differ appreciably for different glasses.

The relaxation time of viscous yield for the value of the viscosity represented in Figure 3 would be about 60 days; it is almost certain therefore that the specimen of this curve was quite unstabilized, and, moreover, that no appreciable change in properties took place during the two or three days over which the load was applied, to judge from the apparently constant slope of BC.

Measurements of viscosity should provide a useful tool for investigations into the process of stabilization, as they can easily be carried out at high temperatures, and the changes which take place in the property of viscosity are relatively much greater than those of any other property. A 25-fold increase in viscosity was observed by Lillie; this corresponds probably to about a 5-fold increase in electrical resistivity, and about 0·5% increase in refractive index. These experiments are, however, greatly complicated by the presence of the delayed elastic effects, which are imitative of a progressive increase in viscosity.

At the present time there have been very few measurements in the range between room temperature and the lower limit of the annealing region of such a kind as to provide accurate information on the property of viscosity (the results of experiments by König (1925) in this range are not easily amenable to analysis since it is not possible to separate the reversible and irreversible parts of the strain from the data given). The measurements so far extend only up to about 10^{18} poises, and it would be of interest to extend them much further. For ordinary inorganic glasses this would probably mean carrying out experiments at or near room temperature; sensitive apparatus and techniques would be required, but a measurement of the viscosity of unstabilized glass at room temperature need not necessarily require a very long period as the delayed elastic effects are over relatively quickly.

There are many analogies between the electrical and mechanical properties of glass. The electrical analogue of the viscosity is the resistivity to direct current, and Ohm's law holds when the electrical analogue of delayed elasticity (the "absorption current") is thoroughly "worked out". The viscosity increases much more rapidly with lowering temperature than does the electrical resistivity, the ratio between the calculated activation energies usually being about 4. It is

of interest that most workers have agreed that the conductivity of ordinary inorganic glasses is nearly all due to the sodium ions. This has been demonstrated experimentally, particularly in the classic experiments of Warburg (1884) and his co-workers. Faraday's laws hold accurately, and apparently a high proportion (75% or more) of the sodium ions present take part. The greater activation energy for viscous flow is clearly due to the fact that the elementary process involves the breaking and reforming of Si–O bonds, which requires much more energy than the passage of the relatively loose sodium ions from one equilibrium position to another.

§ 7. ELASTICITY

The instantaneous elastic strain which occurs on application of a stress will not be discussed here; the value of, say, Young's modulus does not vary appreciably over the range of temperature in which measurements can be made, and does not differ very greatly from the value for a corresponding crystalline material.

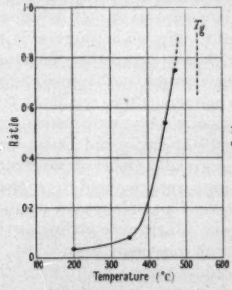


Figure 7. Variation with temperature of the ratio $\frac{\text{Delayed elastic strain}}{\text{Instantaneous elastic strain}}$.

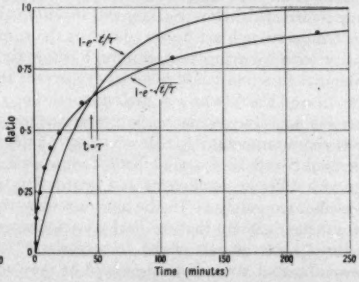


Figure 8. Experimental values of delayed elastic recovery (from Figure 3) as ratio of total, compared with two functions.

Of greater interest are the delayed elastic effects which, as shown in Figure 2, increasingly mask the true viscous flow as we lower the temperature below the annealing or transformation region.

The variation with temperature of the total magnitude of the delayed elastic effect is shown in Figure 7 (Jones 1944) for the same glass as that considered in Figures 2, 3 and 5. It will be seen that this quantity, expressed as the ratio of the delayed part of the elastic strain to the instantaneous part, that is, as c/a, increases rapidly with rising temperature—this is somewhat analogous to the development of "high elasticity" in many organic polymers. At room temperature the ratio has a value lying between 0·1% and 3%, according to the composition (the lowest values being shown by pure silica glass), as shown by many early experimenters, such as Weidmann (1886) and others. At temperatures just below or in the transformation region it becomes increasingly difficult to separate the three parts of the strain, and there is as yet no reliable information about the behaviour here.

As far as can be ascertained from present data, the function $f(t)$ probably has the rather unexpected form $1-\exp\{-\sqrt{(t/\tau)}\}$ (Michelson 1917, Stott 1937, Jones 1944), where τ is a constant. This is illustrated in Figure 8, where the experimental values of the recovery in Figure 3 are plotted against time as fractions

of the total recovery, together with two curves, one representing the function $1-\exp\{-\sqrt{(t/\tau)}\}$ and the other the ordinary exponential function $1-\exp(-t/\tau)$, which is frequently encountered in rheological studies. The same value of τ has been used in each case; where the simple exponential formula is obeyed this constant may be regarded as a relaxation time. Following Alfrey (1948), we shall refer to it as a "retardation time" where it occurs in a function of one of the above types representing delayed elasticity.

It will be seen that the points lie fairly accurately (easily within the experimental error) on the curve of the function $1-\exp\{-\sqrt{(t/\tau)}\}$. As the two curves coincide at the point $(\tau, 1-1/e)$ it is clear that choosing another value for τ would not cause the points to lie any better on the ordinary exponential curve.

The results of Taylor and co-workers (1937), who carried out experiments over a narrow range of temperature in and just below the transformation temperature, differ from the above-mentioned conclusions in two important respects. First, they stated that the magnitude of the quantity c/a was zero at T_g and about 0·5 at a temperature 50° below T_g; by extrapolation they considered that its value at room temperature must be very large as the results indicated an increase in the quantity with lowering temperature. Secondly, it was stated that the simple exponential function could generally represent their results on the development of the delayed elastic strain at each temperature, the value of the retardation time decreasing rapidly with increasing temperature—and so, by extrapolation, very long at room temperature. (It was later (Taylor 1941, Taylor and Doran 1941, Taylor and Noyes 1944) stated that for some glasses a small number of exponential terms with differing coefficients and retardation times were required to represent the results accurately.) In the author's view there is a single reason for these discrepancies, namely that in Taylor's experiments, which were mainly carried out near T_g, a large part of the delayed elastic strain must have been "lost" in the instantaneous strain before it could be measured. It was actually stated that 5 seconds were allowed for the development of the instantaneous part of the strain; of the decay function had been exponential this might not have led to serious error, even near T_g, but in fact this development is very much more rapid in the initial stages than can be represented by an exponential function (at $t=0$ the slope of the function $1-\exp\{-\sqrt{(t/\tau)}\}$ is infinite whilst that of $1-\exp(-t/\tau)$ is $1/\tau$). The experimental identification of the form of this decay function is a good example of the way in which it is possible to bring a particular property into "focus" by working at a suitably chosen temperature.

Taylor considered that the discrepancies between the results obtained by Stott and himself arose from the fact that his own glasses were stabilized, whereas Stott's were not. As already pointed out, the data of Figures 3 and 8 were also obtained on unstabilized glass, and they support Stott's conclusions. The author believes, however, that Taylor's explanation is incorrect and considers it probable that the general form of the function for delayed elasticity (though not the total magnitude or retardation time) would not change fundamentally with the condition with respect to stabilization. The interpretation of the data presented later is based on this point of view.

Other workers (Phillips 1905, Borchard 1935, Murgatroyd and Sykes 1947) who have conducted measurements at room temperature have suggested that the function $f(t)$ is of the form $\log t$. This cannot be correct since its value at $t=0$ is $-\infty$. Moreover, this function does not show the asymptotic approach to a

final value, and all experimenters seem to be agreed that this is the actual behaviour, both in strain and recovery. In fact, where this function has been used, it has failed to represent the experimental results for small or large values of t. Phillips, for instance, considered that t was to be "reckoned from an instant a little before the establishment of the pull". The same is implicit in the statement of results by König (1925). Where the measurements have been continued to high values of t, the formula has again been found inadequate (see for example Murgatroyd and Sykes 1947). Some workers have represented the delayed strains in terms of a function of the form t^m; this again does not represent the asymptotic approach to a final value in strain and recovery.

The variation with temperature of the constant τ has not yet been investigated accurately, but, as pointed out by Stott, its value is surprisingly small at low temperatures, being only minutes or hours at room temperature. Moreover, it apparently does not vary appreciably with temperature, in contrast to the relaxation time for viscous flow which is at least 10^6 times as great at room temperature as it is at the transformation temperature. There is little doubt that the measurements by early workers of delayed elastic strains at ordinary temperatures represented true measurements of the full effect, although the results of Taylor would suggest otherwise.

Very little information yet exists on the variation of the magnitude of the delayed elastic effects with the condition of the glass with respect to stabilization, although the increased effect shown by rapidly cooled glass fibres was reported by Boys in 1887. (Murgatroyd has recently (1948) shown that after "annealing" treatments these effects progressively approach in magnitude those shown by ordinary "massive" glass.) Neither has the form of the decay function been accurately confirmed at a number of temperatures, or for glasses of various compositions. Perhaps the most important gap in the data is, however, that the curve of Figure 7 has not been extended into and above the transformation region. These experiments are difficult because of the fast rates of deflection involved and the preponderating effect of true viscous flow at these temperatures. It should be possible, however, by the use of special techniques, to obtain accurate information above the transformation region. For instance, by applying an alternating stress with differing amplitudes in the positive and negative directions, but with a time-integral of zero, the effect of (Newtonian) viscous flow could be eliminated altogether. The application of high-frequency "square-wave" mechanical stresses, together with suitably developed electrical methods of recording the development of the strain, would also be a powerful method of attack; useful results have already been obtained by such methods applied to the corresponding electrical problem (Whitehead 1946).

Many investigators, such as Birch and Bancroft (1942), Rötger (1941), and others, have studied the "internal friction" of many glasses under ordinary alternating stresses at temperatures which were sufficiently low for the damping to have been mainly due to the development of delayed elastic strains rather than to ordinary viscous flow. An obvious indication that the damping at ordinary temperatures is mainly due to the delayed elastic effect is given by the fact that "lead glasses" have the best "ring" (i.e. least damping) of the common inorganic glasses in spite of the fact that they show the lowest viscosity values in the temperature ranges over which measurements have been made. Probably the most significant result of these investigations is that the maximum dissipation of

energy occurs at a period of vibration which is very much shorter than the retardation time τ. Parallel sets of data do not appear to exist, but Rötger found a maximum internal friction for a fairly typical glass at a period of the order of seconds at room temperature, and of the order of hundredths of a second at about 100°, while τ is usually found to be of the order of minutes or hours, as already indicated.

That the chemical composition has an important effect on the magnitude of the toal delayed elastic effect was shown in the classic investigations of Weidmann (1886). His main conclusion was that glasses containing two alkalis, such as potash and soda, showed the largest effect, and those containing one alkali the smallest effect. This is also true for the well known reversible thermal after-effects* which are responsible for the short-term creeping of the zero and wandering of mercury-in-glass thermometers. As a general rule, a glass with a large delayed elastic effect has a large thermal after-effect, and there is obviously a connection between them; the author has recently (Jones 1948b) attempted to develop a more specific relationship between the two processes.

There is again an electrical analogue to the mechanical after-effect; the corresponding electrical effects have been known for a long time and were studied by Benjamin Franklin, who in 1748 described "residual charges" which he found in discharged Leyden jars. As would be expected, the main feature is that on applying an electric potential difference to glass the current is at first much larger than the final equilibrium value; the initial large current is sometimes called the "absorption current". Under an alternating potential difference there is a power loss, somewhat analogous to that found in paramagnetic materials but due to an electric "power factor", which increases rapidly with temperature until eventually that due to the ordinary electrolytic current predominates. Again, glasses containing lead and barium have low power factors. The influence of the electrical after-effect may also be seen in the observation that the measured resistivity under an alternating potential difference falls off at higher frequences (Robinson 1932); this is presumably due to the fact that the rapid first stages of the "electrical after-effect" are able to play an increased part in carrying the current.

§8. INTERPRETATION OF THE FLOW DATA

(i) *Representation by Mechanical Models*

We shall not consider existing theories of Newtonian viscosity but start from the definition of viscosity expressed in Maxwell's well known equation of visco-elasticity:

$$\frac{dF}{dt} = E\frac{dS}{dt} - \frac{F}{\tau}, \qquad \ldots\ldots(3)$$

where F and S are stress and strain respectively, E is a modulus of elasticity, and τ the relaxation time. If E is the modulus of rigidity, the product $E\tau$ is the coefficient of viscosity as previously stated. This equation leads to a relationship

* The thermal after-effects consist of slow additional changes in volume in the same sense as the instantaneous changes in volume following a variation in temperature. Thus they are analogous to the slow volume changes of stabilization. However, in glass at or near ordinary temperatures they represent only minor changes in a structure which is essentially still unstabilized, and they are very much smaller in magnitude than the configurational contributions to the changes in volume which would be observed if fully stabilized glass could be obtained at the same temperatures. Their equilibrium times are of course very much shorter than that of the stabilization process, being about the same as those of the delayed elastic effects. They can perhaps be regarded as local or "permitted" attempts at stabilization.

of the following type for the development of the strain under a constant applied stress:

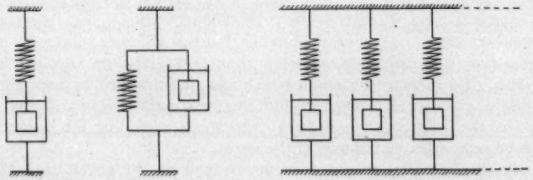
$$\text{Strain} = a + bt \qquad \ldots\ldots(4)$$

where a and b are constants in any given arrangement and $a/b = \tau$. Thus it does not include any representation of delayed elasticity but only serves to give exact formal definitions of the modulus of (instantaneous) elasticity, the coefficient of viscosity and relaxation time. However, as we shall see later, it is possible by postulating a distribution of viscosities or relaxation times to account for the existence, and to some extent for the form, of the delayed elastic effect.

Another solution of Maxwell's equation must be mentioned here on account of its importance in the theory of annealing. If the strain is kept constant by some means the stress disappears according to the equation

$$\text{Stress} = F_0 \exp(-t/\tau), \qquad \ldots\ldots(5)$$

where F_0 is the value of the stress at time $t = 0$. The release of stress according to this function is often known as "Maxwellian"; as already stated, the relaxation



(*a*) Maxwell element. (*b*) Voigt element. (*c*) A generalized Maxwell model.

Figure 9.

time is the time taken for the stress to "relax" to $1/e$ of its original value through viscous flow, when the strain is kept constant. It is important to note that this equation does not itself account for delayed elasticity, as wrongly assumed by some authors.*

For the purpose of interpretation of the data already summarized it is very convenient to make use of the so-called mechanical "models" which were introduced many years ago by J. J. Thomson, Wiechert (1893) and others, and which have recently received much attention because of the growing importance of the mechanical properties of organic high polymers. These models are useful as a "shorthand" for the representation of particular properties and make the analysis of any problem much clearer and therefore simpler.

In Figure 9, model (*a*), consisting of an elastic element and a viscous element (or spring and "dash-pot") in series, is generally known as a "Maxwell element", since its extension under load is that represented by Maxwell's equation. It should be made clear that the representation of elastic properties by means of a spring is not intended to mean that the particular material under consideration possesses ideal solid, that is, perfectly crystalline regions; it simply represents

* Maxwell's only suggestion as to the cause of the delayed elastic effect referred to metal wires in torsion and required a dependence of "viscosity" on stress. In the case of rapidly cooled glass fibres there might actually be a lower viscosity in the outer layers and a mechanism similar to that suggested by Maxwell might play a part. This has, however, no bearing on the origin of the ordinary delayed elastic effects in glass, for which the viscosity is independent of the stress.

the fact that the instantaneous response of the structure to an applied stress is truly elastic. As we have already seen in the curves of Figure 2, a material which shows viscous flow can also show instantaneous elasticity and would obviously show roughly the same instantaneous elasticity even at much higher temperatures, where the viscosity would be much lower and the viscous flow much faster. The instantaneous extension of the structure is in fact nearly similar to that which would be shown by a true solid lattice.

Another model, generally known as a "Voigt element" (Voigt 1928), is illustrated in Figure 9(*b*). The equation of motion under constant stress is of the form

$$\text{Strain} = c\{1 - \exp(-t/\tau)\} \qquad \ldots\ldots(6)$$

where τ is now the constant which we have called the "retardation time". Thus this model exhibits delayed elasticity only. Many organic substances do appear to act roughly in this way although the equations are generally not of such simple form.

By combining Maxwell and Voigt elements in series we have a model which obeys an equation of the type

$$\text{Strain} = a + bt + c\{1 - \exp(-t/\tau)\} \qquad \ldots\ldots(7)$$

under constant stress. (It should be noted that two different relaxation times—or a relaxation time and a retardation time—may now be involved in this expression.) To an approximation, which can be improved by introducing in series additional Voigt elements with differing retardation times and coefficients, this imitates the behaviour of most real substances, including the instantaneous and delayed recoveries on removal of the stress.

Exactly similar equations are however obeyed by the generalized Maxwell model shown in Figure 9(*c*) in which Maxwell elements with differing characteristics are connected in parallel, and this model is probably more instructive for our purpose. As early as 1893, Wiechert computed strain–time curves for such models consisting of a small number of elements with a Gaussian distribution of $\log \tau$, and much theoretical work of this type has since been carried out by Bloch (1945), Burgers (1935), Kuhn (1939), and others. Simha (1942) has considered the case of a body in which a continuous distribution or "spectrum" of relaxation times exists, and has shown how, by employing Fourier transformations, a theoretical distribution function may be calculated for any given "creep curve", and vice versa. Any creep curve may in fact be analysed either in terms of the generalized Maxwell model with a distribution of relaxation times or in terms of the model previously mentioned consisting of a single Maxwell element in series with a number of Voigt elements with a distribution of retardation times. The same applies to, say, a curve representing stress release against time, or a curve giving the variation with frequency of the internal friction under oscillating stress. These problems, which are mathematically analogous to problems frequently arising in the theory of alternating current networks containing resistances and capacities, have recently been comprehensively reviewed by Alfrey and Doty (1945, see also Alfrey 1948).

As shown by Bennewitz and Rötger (1939), the behaviour under constant stress of the generalized Maxwell model consisting of N Maxwell elements in parallel is given by an equation of the following type:

$$\text{Strain} = a + bt + \Sigma c_i\{1 - \exp(-t/\lambda_i)\}, \qquad \ldots\ldots(8)$$

where i has values 1 to $N-1$. The delayed elastic effects are thus represented by $N-1$ exponential terms with differing retardation times and coefficients (we may remark that a number of experimenters have represented their results as the sum of a small number of such exponential terms). A similar equation would of course be obeyed by a model of the other type considered if it contained $N-1$ Voigt elements in series with a single Maxwell element. The evaluation of the various coefficients and the retardation times λ_i (which are not simply the individual values of relaxation time, say $\tau_1, \tau_2 \ldots$ of the constituent elements) for the generalized Maxwell model is cumbersome, but many properties of the model may be demonstrated by considering the case $N=2$, when the equation reduces to the following for constant stress, if we make the simplifying assumption that the elastic properties and dimensions, or "form factors", of the two elements are identical:

$$\text{Strain} = a\left[1+\frac{2}{\tau_1+\tau_2}t+\left(\frac{\tau_1-\tau_2}{\tau_1+\tau_2}\right)^2\left\{1-\exp\left\{-\frac{1}{2}\left(\frac{1}{\tau_1}+\frac{1}{\tau_2}\right)t\right\}\right\}\right]. \quad \ldots\ldots(9)$$

The following implications of the above expression are of interest: (*a*) the "viscosity" of the composite body is equal to the arithmetic mean of the viscosities in the constituents, as the relaxation time of viscous flow (a/b in the notation of the previous equations) is equal to $(\tau_1+\tau_2)/2$; (*b*) the delayed elastic strain is always in the same direction as the instantaneous strain, since its coefficient is always positive; (*c*) the total magnitude of the delayed elastic effect is a rapidly increasing function of the "spread" of relaxation times about a given mean (as originally suggested by Kuhn on more general grounds); (*d*) the retardation time of the delayed elastic effect is equal to the harmonic mean of the relaxation times of the constituents; it is thus smaller than the viscous relaxation time of the compound body.

Qualitatively, therefore, many features of the experimental data already mentioned can be represented by this simple model if we simply assume that the "spread" in the relaxation times increases with increasing temperature. A possible explanation is suggested here also (see (*d*)) for the fact that the retardation times of the delayed elastic effects are shorter than the relaxation times of viscous relaxation; this point is, however, discussed again later.

By considering a larger number of parallel Maxwell elements we can represent delayed elastic effects (such as those found with glass) which are more rapid at first, and slower later, than would be represented by an exponential function. The decay function for delayed elasticity will now consist of a large number of terms with complicated coefficients and retardation times; we do not here evaluate any more elaborate case but simply consider a model for which the distribution curve of relaxation times (or log (relaxation times)) has a long "tail" on each side, that is, a small number of elements with very high and very low values of relaxation times. It is first convenient to explain in familiar terms how the generalized model of Figure 9 (*c*) leads to a delayed elastic effect. For convenience we consider all the springs identical. Then when the load is applied all these springs are immediately extended to the positions which they would occupy in the absence of the "dash-pots", and the load is distributed equally between them. However, the springs in the elements containing less viscous fluid immediately begin to contract again because of viscous flow under the force transmitted by the spring. A smaller fraction of the load is now taken by these springs and the remaining springs consequently extend further, giving the delayed elastic effect.

At equilibrium, that is when the flow is purely viscous, the forces in the elements are in the ratios of their viscosities as the velocity is everywhere the same (it follows from this that the viscosity of the body is equal to the arithmetic mean of the individual viscosities, as already stated for the case $N=2$). In a similar way, an explanation can be given for the delayed elastic recovery.

It will be clear that the effect of a "tail" at low relaxation times in the distribution curve must be to cause a very rapid development of the delayed elastic effect immediately after applying the load; similarly, the "tail" at high relaxation times slows down the later stages of this development and of the recovery. Thus we can account qualitatively for the form of the departure of the decay function from the simple exponential formula. We can probably explain similarly the fact that the internal friction under oscillating stress is a maximum at a period of oscillation much shorter than the usual values of τ.

In the same way the form of the function for delayed elasticity can be represented by a distribution of series Voigt elements. The author has submitted the function $1-\exp\{-\sqrt{(t/\tau)}\}$ to analysis in these terms; this leads to a nearly symmetrical distribution of log (retardation time) about a maximum at $\log\tau$, again with "tails" at high and low values which are actually asymptotic in their approach to the axis of log (retardation time). In fact, whichever model is adopted, this function implies a wide distribution of relaxation times with tails at both high and low values of log (retardation time). The same would clearly apply to more complex models or combinations of models.

(ii) *Significance of a Distribution of Relaxation Times*

We must now consider what physical meaning can be attached to the distribution of retardation times among series Voigt elements or relaxation times among parallel Maxwell elements in the case of the material and experimental data under discussion. It must be stated that the large amount of work which has been carried out on the properties of high polymers has not yet yielded any clear picture of the general relations between structures and flow properties. In the case of linear high polymers, however, it is generally held that the delayed elasticity can be attributed to the uncoiling of the chain molecules, and the true viscous flow to the relative movement of whole molecules (see for instance Alfrey 1948). The introduction of several series Voigt elements is then an indication that there are several processes with differing retardation times, such as the uncoiling of short segments of the molecules and the uncoiling of long segments. The different Voigt elements represent different mechanisms or modes of response to a given stress (since in this model each element is under the same stress); since the total response is limited in each case and takes place by diffusion processes similar to those of ordinary viscous flow, the Voigt element is a reasonable representation. Taylor actually attributed the small number of separate exponential terms in the functions he used to represent delayed elasticity in glass as due to the action of individual ion types; this is almost certainly an over-simplification, particularly as we have seen that a much wider distribution of retardation times is actually required to represent the experimental results. (The author believes that Smekal's general explanation of these and related phenomena (1938, 1939, 1940) is also over-simplified, and suffers from too much reliance on Taylor's experimental data.)

Another physical conception which can reasonably be represented by the Voigt element is the gel structure, in which more or less solid regions completely surround liquid regions; the total strain under stress is limited by the elasticity of the solid part, but its development is "damped" by the presence of the liquid (see for instance Rötger 1941, Fröhlich and Sack 1946).

It seems to the present author that the clearest interpretation of the flow properties of glass is given if we consider the structure in terms of the generalized Maxwell model consisting of parallel Maxwell elements, that is, as being truly amorphous or "liquid", but with a distribution of relaxation times. Two possibilities now suggest themselves as to the significance of this distribution.

The first, and perhaps the more obvious, possibility is that we have to consider a real spatial distribution of viscosity such as might arise from the presence of chemical inhomogeneities. It is well known that inhomogeneities of composition of colloidal dimensions do exist in ordinary glasses, and that the "spread" in the composition can be progressively reduced by repeated grinding and melting (Tooley and Tiede 1944). There is, unfortunately, no evidence as to whether specially prepared homogeneous glass shows smaller delayed elastic effects. Pure silica is of course a one-component glass and, therefore, chemically homogeneous, but it is very difficult to prepare *physically* homogeneous fused silica on account of its high viscosity at the melting temperature. Moreover, its transformation region is so high (~1,200°) compared with that of ordinary glasses that it is unsafe to draw any conclusions from the fact that the smallest delayed elastic effects are shown by silica. It would be of great interest to extend experiments on the delayed elastic properties of silica and of other one-component glasses, such as glycerol, and also to investigate further ordinary glasses with special reference to the influence of the degree of homogeneity.

The second, and more fundamental, possibility is that the distribution of relaxation times postulated may in effect be simply the distribution of potential trough depths or barrier heights which exists in the structure. Ordinary theories of viscosity involve the concept of an "average activation energy" which an atom has to acquire before it can surmount a potential energy barrier and enter a new equilibrium position. In reality there is of course a wide distribution of "activation energies".

Immediately after application of a stress the first migrations to occur will be those which require the lowest activation energies, and the resulting rearrangements will, on the average, lead to an alteration of the structure in the sense that there will now be fewer migrations possible of low activation energy. After removal of the stress the structure slowly returns to its original condition since this is a condition of minimum free energy.

Thus the origin of the delayed elastic effects might be a purely statistical matter; it would seem that development of theories of viscosity so as to take account of the distribution of activation energies might lead to a representation of these effects. Smith (1948) has recently put forward a theory of "transient creep" in metals which appears to be based on similar considerations.

There is of course no sharp distinction between the two mechanisms discussed above; inhomogeneities of composition may sometimes be very small and the groups whose movements are responsible for viscous flow may be appreciably larger than atomic in size. The following additional remarks concerning the interpretation of the data probably apply equally whichever mechanism is

considered. We have already seen that the "spread" of relaxation times must increase with increasing temperature to account for the increased delayed elastic effect; we cannot say much more at present than that this seems a reasonable picture, and one which is in accord with the familiar conceptions of statistical mechanics. Another feature of this method of representation is that the general form of the function for delayed elasticity could remain the same in spite of large changes in its total magnitude. The relation between "spread" of relaxation time and magnitude of the delayed elastic effect also offers an explanation for the fact that larger effects are shown by glasses containing more than one type of alkali ion, particularly when, as in the case of sodium and potassium, these are of differing size and coordination number towards oxygen. It is perhaps worth mentioning that the curves of Figure 2 provide sufficient justification for the assumption of identical elastic properties in the separate Maxwell elements.

We must consider briefly what physical significance is to be attached to the method of "coupling" involved in the model of Figure 9(*c*), which we have mainly used in this discussion. In the model, the separate elements are imagined to be coupled to rigid bars in such a way that the extension is always the same for all elements at a given moment, the extension of a single element being restricted by the presence of all the other elements. If we consider the origin of the delayed elastic effect to be a distribution of chemical composition, the parallelism lies in the fact that the regions of differing composition are in close contact with each other. If the origin of the effect is the distribution of activation energies we can only say that the processes governed by a particular value of activation energy or relaxation time are affected, or restricted, by all other similar processes governed by differing values.

(iii) *Activation Energy of Viscous Flow*

Ward (1937) has considered the question of the variation of the activation energy of viscous flow for liquids investigated over a wide range of temperature, on the basis of Bernal's theory of liquid structure. He attributes the decrease in the activation energy with increasing temperature, shown for instance by boric oxide glass, to a progressive increase in the average degree of coordination of a given atom or molecule by other atoms or molecules, that is, to a reduction in structural character. This was the behaviour expected by Bernal for liquids with directional bonds, of which the inorganic glasses with their highly polarized ionic bonds are examples, and actually observed by Bernal and Fowler (1933) in X-ray studies on water.

Ward suggested other ways in which changing temperature might affect the activation energy by changing the type of force present. Thus at higher temperatures weakly bound molecules (or other aggregates persisting for times long compared with the Maxwell relaxation time) might become dissociated, or new types of relative rotation of groups might become possible. The former mechanism was expected to lead to an increase in the value of the activation energy B with increasing temperature, and Stanworth (1937) suggested that the converse process—the formation of "clusters"—might account for the reduced values of B below the transformation region. Reduced values of the calculated activation energy are undoubtedly shown by *unstabilized* glass below the transformation region and, as we shall see, there is independent evidence for the occurrence of local configurational adjustments below the transformation region of a type which might lead to "clustering".

(iv) *Further Evidence for Configurational Adjustments below the Transformation Region*

As has been stated, the retardation time of the delayed elastic process is much shorter than the relaxation time of viscous flow for the case of unstabilized glass at ordinary temperatures, though probably of the same order in and just below the transformation region. We do not know whether it continues to be the shorter of the two relaxation times in the transformation region, although, as we have seen, this is suggested by the computation of the simple generalized Maxwell model containing two parallel elements. At ordinary temperatures, however, the difference is so great that we are again forced to consider whether minor configurational changes might not be able to occur during cooling from the temperature in the transformation region at which the main features of the configuration are "frozen-in". For if a given model is assumed to represent the frozen-in configuration, the effect of lowering the temperature would be to increase the values of log (viscosity) in the separate components by factors of about the same order of magnitude. The two relaxation times mentioned above would therefore increase, very approximately, in the same ratio. The fact that they do not, suggests that the model itself is altered, that is, that local structural adjustments are able to take place below the transformation region. Since glass below the transformation region is in an unstable state, with a tendency to contract to a smaller volume, and also possesses a wide distribution of relaxation times, it would not be surprising if such changes could occur at favourable points. (The somewhat related suggestion that some of the changes which can take place during stabilization might be appreciably more rapid than others has been made by Tool (1946).) Gurney (1947) has suggested that similar rearrangements or local contractions might be responsible for the "Griffith flaws" which are postulated to account for the low mechanical strength of glass.

The assumption of a fixed configuration below the transformation region is therefore almost certainly only a first approximation, sufficient to explain broadly the main transformation phenomena, such as the temperature-variation of the volume and thermodynamic properties, but insufficient to account for certain other aspects of the behaviour. The flow properties are especially sensitive to second order or local changes in structure since they depend on individual migratory processes, and the property of mechanical strength is the most markedly "structure-sensitive" of all properties.

The type of configurational change which can occur can probably be best described by the term "clustering", already used. Thus relaxation at regions or points of low viscosity would permit other regions to contract very slightly towards their equilibrium densities even below the transformation region. As will be clear from the preceding discussion of the flow data, we can hardly give any estimate as yet of the sizes of these clusters. Also, the question of whether these configurational adjustments are entirely reversible with changing temperature below a given "transformation temperature" is as yet quite unanswerable from the data, but this seems to the author unlikely.

(v) *Other Features of Interest*

On the basis of the foregoing general interpretation of the data, predictions may be made on a number of topics not yet covered by experiment. We should, for instance, probably expect the retardation time of the delayed elastic effect to increase rapidly during a "stabilization" treatment, though the form of the decay

function might remain unaltered. The total magnitude of the effect would decrease (as already shown by Murgatroyd for rapidly cooled fibres), so that the variation with temperature of this quantity for completely stabilized glass would continue to be in the same sense as for ordinary unstabilized glass, that is, it would decrease with decreasing temperature; this suggestion differs from the expectations of Taylor and co-workers (1937). We would not expect any changes more fundamental than changes in magnitude to be shown in any of the flow properties if it proved possible to obtain glass completely stabilized at temperatures down to room temperature. Thus there would undoubtedly be a finite value of the viscosity, and the yield point would still be at zero stress. The average activation energy for viscous flow would probably continue to increase with falling temperature.

It is perhaps more difficult to predict what is likely to be found above the transformation region, but it seems very likely that the value of the ratio c/a will continue to increase with increasing temperature, though of course the retardation times of the delayed elastic effects will now be very short.

Mention must be made of a somewhat different type of visco-elastic behaviour affecting the volume (see for instance Frenkel 1946). The delayed elastic decrease in volume under a uniform compressive stress is considered as a slow approach to equilibrium volume under changed pressure conditions, analogous to the slow change in any property which occurs below the transformation region under changed temperature conditions. We have considered the ordinary delayed elastic effect as being related to the viscosity, that is, as being a reaction to shear stresses. This is probably justified at temperatures below the transformation region, where the retardation time of the delayed elastic effect and the equilibrium time of the ordinary stabilization process differ by several orders of magnitude. In any case there seems as yet to be no experimental evidence about the delayed elastic behaviour of glass subjected to uniform compression.

The foregoing discussion has necessarily been somewhat sketchy and mainly exploratory. In spite of the fundamental importance of many of these problems there are many serious gaps in the experimental data which will have to be filled before much progress can be made in explanation and interpretation. It seems to the author that further experiments in this field are likely to be of great value in the development of the theory of the liquid state, and, of course, of the theory of liquid viscosity. They promise also to be of more general interest to the theoretical rheologist, since the behaviour of the components of any model considered can be assumed to be relatively simple in the case of glass, in contrast to other materials, where non-Hookeian elasticity, non-Newtonian viscosity, non-reversible transient creep, yield points, work-hardening and other complicating phenomena are encountered.

§9. PRACTICAL ASPECTS OF THE DATA IN THE TRANSFORMATION REGION

The most important practical application of the data in the transformation region is in the annealing of glass. One of the main functions of annealing is the removal of mechanical stresses between adjacent layers, since these stresses lead to birefringence and can also cause spontaneous fracture if they amount to $\frac{1}{4}$ or $\frac{1}{8}$ of the short-term breaking strength at any point. For ordinary commercial articles, permanent stresses are therefore kept down to about 1/20 of the breaking

strength, except in special cases where permanent stresses introduced under carefully controlled conditions can be desirable. In "toughened" glass, for instance, large surface compression stresses are introduced by deliberately accelerating the cooling of the surfaces in a blast of air or by immersion in a molten salt mixture.

The way in which rapid cooling leads to surface compressive stresses and internal tensile stresses has been very adequately explained by Adams and Williamson (1920), and only a brief summary can be given here. We consider a sheet of glass cooling at a uniform and fairly rapid rate from above the transformation or annealing region. An approximately parabolic temperature gradient will exist across the thickness of the specimen, the surfaces being at a lower temperature than the interior. No stresses can exist in the specimen above the annealing region, and as an approximation we can say that there will therefore be no stresses in the specimen when the surfaces reach room temperature. Now, however, the temperature gradient must disappear as the interior must also reach room temperature, and as the viscosity is now too high for any further relaxation of stresses, the contraction of the interior leads to permanent stresses in the sense described above.

Detailed consideration of this process shows that the critical region through which cooling must be slow (if permanent stresses are to be avoided) is the lower part of the annealing region. The normal process of relieving stresses during annealing therefore consists of heating the specimen into the annealing region and maintaining it there until viscous relaxation of the existing stresses is complete, and then cooling slowly through the critical region. After passing through this region the remainder of the cooling may be carried out as rapidly as is possible without causing fracture as a result of the temporary stresses introduced.

The first theoretical study of the process was made by Twyman (1917), who assumed that the relaxation of stress at a constant temperature in the annealing region followed the exponential law put forward by Maxwell (equation 5). The equilibrium time for the release of stresses should therefore be the relaxation time calculated from the rigidity and viscosity values. (It is not always realized that the permanent stresses in glass must always consist of equal and opposite stresses between adjacent layers; it is this fact that keeps the strain at any point constant, as required by equation (5). An isolated piece of glass in which the stresses are all of the same sign is an impossibility.)

It was soon found, however, that the simple exponential formula could not nearly represent the rate of release of stress in glass at constant temperature, the actual rate of release immediately after raising the glass to this temperature being much more rapid than would be expected from the known value of the viscosity. For practical purposes empirical formulae, such as that proposed by Adams and Williamson, have been used with fair success (though they usually fail to represent the behaviour during the period immediately after the start of the process). Many explanations were proposed to account for the apparent anomaly, some of these being untenable from the point of view of dimensional analysis. It was not until Littleton (1933) and Lillie (1936) pointed out that the viscosity generally changed continuously in a specimen heated into the annealing range, and Lillie carried out parallel sets of experiments on the rate of annealing and the rate of change of viscosity at constant temperature, that any explanation was possible. Tool later (1945) derived an equation which represented experimental data reasonably well on the basis of a continuously changing "fictive" temperature.

However, both these authors neglected to consider the delayed elastic effects which, as recently shown by Jones (1947, 1948a), must have a very significant effect on the rate of release of stress, greatly accelerating it in its early stages and retarding it later. The reason for the effect is obvious if we consider that the value of the ratio c/a is about 0·75 or more in the annealing region but only about 0·01 at room temperature, so that in the complete absence of viscous flow, raising the temperature to the annealing region would eventually reduce permanent stresses by a factor $1/(1+0{\cdot}75)$, that is, to about 0·6 of their original value. When completely annealed, of course, no delayed elastic strains remain, so that at some stage they begin to recover, and the basic mechanism for the release of stress is still viscous flow. Any formal theory of the complete flow properties which included the delayed elastic effects in terms of adequate models or equations would of course also represent the rate of release of stress correctly (see for instance Mack 1946, Alfrey 1948). The stress-release function actually provides the most convenient method for determining the distribution of relaxation times of the generalized Maxwell model. Experimental amplification of the above statements would however be complicated by uncertainty about the delayed elastic effects frozen-in during the previous cooling.

The reason why delayed elastic effects have not usually been considered must be that they are so imitative of a progressive increase in viscosity that their effect in any investigation would appear also in the parallel experiments on the change of viscosity. However, it can easily be shown that this must lead to anomalies, and it is of interest that both Lillie and Tool stated that anomalies persisted in the interpretation of their results. Lillie attributed these to an "elastic reaction", but obviously did not realize the major part which could be played by elasticity, nor that each "reaction" must later be reversed. Tool incorrectly attributed them to "plasticity", or a dependence of the viscosity on the value of the stress.

Finally, it must be mentioned that the relaxation of stresses is not the only function of annealing, and in the case of the highest quality optical glass it is actually of only secondary importance. A useful review of this subject has been given by Hampton (1942); the important criterion is that throughout the whole volume of a piece of such glass the material must have the same refractive index, and this means that the temperature gradients in the glass during cooling through the transformation region, where the main freezing-in of properties occurs, must be as small as possible. (Of course, inhomogeneities of composition must be absent also, but this must be dealt with during the original melting of the glass.)

§ 10. PRACTICAL ASPECTS AT ORDINARY TEMPERATURES

Though it is impossible from present data to state in exact quantitative terms what degree of rigidity or long-term permanence in shape can be expected at ordinary temperatures from a specimen of glass given a particular annealing treatment, it may be useful to summarize here the well known facts relating to "creep" or "loss of figure" in glass articles by reference to the particular mechanisms (stabilization, ordinary viscous flow, or delayed elasticity) responsible in each case.

First we mention those effects which are probably due to slow progressive stabilization. The most familiar of these is the slow irreversible creeping-up or "secular rise" of the ice-point in mercury-in-glass thermometers due to the progressive contraction of the material of the bulb (this is quite distinguishable from the short-term reversible thermal after-effects also observed in thermometers).

With unsuitable types of glass this effect can cause a change of zero corresponding to 10° or 20° over a period of years, but normally it is of course much less. The "loss of figure" sometimes shown by pieces of optical glass is also probably due to the same cause, but one would expect this to be serious only if the "frozen-in" configuration had varied from point to point in the specimen at the time of figuring. Different parts of the material might then change in refractive index at slightly different rates, so that "loss of figure" would occur. The slight changes in shape which could also occur if the density changed at different points at appreciably different rates might also have some influence; the slow change in size alone which would occur if the glass had been in the same condition throughout during cooling would hardly be expected to cause loss of figure.

The "secular rise" of the ice-point in thermometers can be reduced to relatively unimportant proportions by heating the thermometer for a long period at a temperature somewhat below the normal annealing range. The purpose of this treatment, which is also sometimes referred to as an "annealing" treatment, differs again from the two functions of the annealing processes already mentioned. It represents an attempt to lower the configurational temperature of the glass as far as possible by maintaining it at a temperature sufficiently high for the rate of stabilization to be appreciable, but sufficiently low for the final configuration to be nearly stable at ordinary temperatures.

Few systematic investigations have been carried out into these very slow changes in properties which occur at ordinary temperatures. Winter (1943) observed changes in the refractive index persisting over 16 months, and Joule observed the behaviour of a mercury-in-glass thermometer for over 40 years, during which time the whole rise in the ice-point was less than 1° F. The observed values of the rise could be fairly well represented by a simple exponential function with an equilibrium time of 15·3 years. It seems, however, that the real equilibrium time of the complete stabilization process in ordinary glasses at ordinary temperatures must be much longer than 1 or 2 or even 15 years, and that the apparent equilibrium usually noted is an indication that the viscosity has increased so much that the rate of change is very much slower. As recognized by Tool (1946), one could not in this case expect the whole behaviour to be represented by a simple exponential function; recent experiments by Collyer (1947) also support this suggestion.

Effects of "creep" due to purely viscous flow are probably more rarely met with. When such flow is due to permanent stresses it should not normally cause distortion unless the properties of the material vary from point to point, and glass is rarely used under externally applied permanent load in such a way as to enable very small irreversible distortions to be observed. Doubt has even been expressed as to whether any viscous flow could be detected in, say, soda-lime-silica glasses at room temperature; it has been variously suggested that the viscosity of ordinary glass here must be between 10^{40} and 10^{70} poises, which would be practically impossible to measure. (These estimates have sometimes been based on the relationship which exists between the viscosity and electrical resistivity of glass at higher temperatures. However, indications of the occurrence of local changes in structure at lower temperatures have already been mentioned, and it is known that the electrical conductivity is entirely due to the sodium ion; in view of these complicating factors, such estimates should be treated with reserve.) The curve of Figure 5 however suggests that the viscosity of an ordinary annealed glass of

typical composition at ordinary temperatures may be between 10^{20} and 10^{23} poises; we do not consider "stabilized" glass at ordinary temperatures as this is probably never obtained in practice. The relaxation times corresponding to these values of viscosity are of the order of 20 and 20,000 years respectively.* Since it appears that delayed elastic effects are always rapidly worked out, it should be possible by accurate measurements of a high grade to obtain a true value of viscosity in a period not exceeding weeks or months. It is not likely however that a roughly set up experiment, or a chance observation, would show any irreversible effect even over months or years. Familiar reports of glass tubes sagging noticeably under their own weight probably arise from the fact that tubes, if originally slightly curved, will settle into the most stable positions on their supports.

In practice the effects of viscous flow are probably most often encountered when a badly annealed piece of glass, with viscosity lower than that of a well annealed specimen, is observed to lose its permanent stresses or "strain" slowly over a period of years, as has often been reported. There is another common manifestation of flow on a small scale in the fact that diamond cuts on glass are known to "age", that is, the sheets become more difficult to break if time is allowed after cutting.

Glass specimens which have been very rapidly chilled are of course much more likely for various reasons to show viscous flow at ordinary temperatures. Murgatroyd (1944) has reported that in the thinnest glass fibres it is possible to induce a "permanent set" in a few days at room temperature, or in a few hours at 100°. These fibres, having been very rapidly cooled during drawing, would have abnormally low viscosity; it is also possible to apply much higher stresses to fine fibres than to specimens of "massive" glass. There is evidence that thin fibres are extremely non-isotropic in their properties; for instance, lower values of viscosity or greater delayed elastic effects are sometimes observed in the case of torsional stresses.

Many reports of "creep" observed in glass must be due to the delayed elastic effects. These are of course reversible and quite small, so that it should always be possible to decide whether they are responsible for a particular case of creep.

To summarize, we may say that in rapidly cooled glass, in which permanent stresses and variation of properties may exist, each of the above-mentioned effects can contribute to nearly all observed forms of "creep". In normally annealed glass they are of magnitudes which should be measurable under careful investigation; in practice they are often noticed after long periods but escape detection when looked for over short periods. Careful annealing and choice of glass provide the somewhat empirical but practically successful answer to most of the problems met with in the use of glass as a rigid material.

ACKNOWLEDGMENTS

The author has had the advantage of helpful discussions with Professor F. Simon and with Dr. A. von Engel on many of the subjects dealt with in this paper. He is grateful also to Mr. R. O. Davies for assistance.

* It is of interest that observations by Ghering and Green (1945) of the apparent permanence of the stresses in glass "strain discs", if interpreted mathematically in terms of their stated limits of certainty, only mean that the relaxation time for their glass was at least greater than about 100 years.

REFERENCES

ADAMS, L. H., and WILLIAMSON, E. D., 1920, *J. Franklin Inst.*, **190**, 597, 835.

ALFREY, T., 1948, *Mechanical Behavior of High Polymers* (New York : Interscience Publishers).

ALFREY, T., and DOTY, P., 1945, *J. Appl. Phys.*, **16**, 700.

ANDRADE, E. N. DA C., 1934, *Phil. Mag.*, **17**, 497.

BECKER, 1938, *Ann. Phys., Lpz.*, **32**, 128.

BENNEWITZ, K., and RÖTGER, H., 1939, *Phys. Z.*, **40**, 416.

BERNAL, J. D., 1937, *Trans. Faraday Soc.*, **33**, 27.

BERNAL, J. D., and FOWLER, R. H., 1933, *J. Chem. Phys.*, **1**, 515.

BIRCH, F., and BANCROFT, D., 1942, *Amer. J. Sci.*, **240**, 457.

BLOCH, A., 1945, *J. Instn. Elect. Engrs.*, **92**, Pt. I, 157.

BORCHARD, K. H., 1935, *Glastech. Ber.*, **13**, 114.

BOYS, C. V., 1887, *Proc. Phys. Soc.*, **9**, 8.

BURGERS, J. M., 1935, *Kon. Akad. Wet., Verhand.*, **15** (No. 3).

COLLYER, P. W., 1947, *J. Amer. Ceram. Soc.*, **30**, 338.

DALE, A. E., and STANWORTH, J. E., 1945, *J. Soc. Glass Tech.*, **29**, 414.

EYRING, H., 1936, *J. Chem. Phys.*, **4**, 283.

FRENKEL, J., 1946, *Kinetic Theory of Liquids* (Oxford : University Press).

FRÖHLICH, H., and SACK, R., 1946, *Proc. Roy. Soc.* A, **185**, 415.

GHERING, L. G., and GREEN, T. D., 1945, *J. Amer. Ceram. Soc.*, **28**, 288.

GURNEY, C., 1947, *Proc. Phys. Soc.*, **59**, 169.

HAMPTON, W. M., 1942, *Proc. Phys. Soc.*, **54**, 391.

JONES, G. O., 1944, *J. Soc. Glass Tech.*, **28**, 432 ; 1947, *Ibid.*, **31**, 218; 1948 a, *Nature, Lond.*, **161**, 813 ; 1948 b, *Proc. Phys. Soc.*, **61**, 320.

KÖNIG, H., 1925, *Phys. Z.*, **26**, 797.

KUHN, W., 1939, *Z. Phys.* B, **42**, 1.

LILLIE, H. R., 1930, *Phys. Rev.*, **36**, 347 ; 1931, *J. Amer. Ceram. Soc.*, **14**, 502 ; 1933, *Ibid.*, **16**, 619 ; 1934, *Ibid.*, **17**, 43 ; 1936, *Ibid.*, **19**, 45.

LITTLETON, J. T., 1933, *Ind. Eng. Chem.*, **25**, 748.

LITTLETON, J. T., and WETMORE, W. L., 1936, *J. Amer. Ceram. Soc.*, **19**, 243.

MACK, C., 1946, *J. Appl. Phys.*, **17**, 1086, 1093, 1101.

MAXWELL, J. C., 1867, *Phil. Trans.*, **157**, 49.

MICHELSON, A. A., 1917, *J. Geol.*, **25**, 405.

MOREY, G. W., 1938, *The Properties of Glass* (New York : Reinhold).

MOREY, G. W., and MERWIN, H. E., 1932, *J. Opt. Soc. Amer.*, **22**, 632.

MOTT, N. F., and GURNEY, R. W., 1939, *Rep. Prog. Phys.*, **5**, 46. (London : Physical Society).

MURGATROYD, J. B., 1944, *J. Soc. Glass Tech.*, **28**, 368 ; 1948, *Ibid.*, in the press.

MURGATROYD, J. B., and SYKES, R. F. R., 1947, *J. Soc. Glass Tech.*, **31**, 36.

OBLAD, A. G., and NEWTON, R. F., 1937, *J. Amer. Chem. Soc.*, **59**, 2495.

PARKS, G. S., and SPAGHT, M. E., 1935, *Physics*, **6**, 69.

PHILLIPS, P., 1905, *Phil. Mag.*, **225**, 520.

ROBINSON, D. M., 1932, *Physics*, **2**, 52.

RÖTGER, H., 1941, *Glastech. Ber.*, **19**, 192.

SIMHA, R., 1942, *J. Appl. Phys.*, **13**, 201.

SIMON, F., 1927, *Z. Phys.*, **41**, 806 ; 1930, *Ergebn. exakt. Naturw.*, **9**, 248 ; 1931, *Z. anorg. Chem.*, **203**, 219.

SMEKAL, A., 1938, *Glastech. Ber.*, **16**, 198 ; 1939, *Z. Phys. Chem.*, **44**, 286 ; 1940, *Ibid.*, **48**, 114 ; 1947, *FIAT Review of German Science, Physics of Solids*, Part 1.

SMITH, C. L., 1948, *Proc. Phys. Soc.*, **61**, 201.

STANWORTH, J. E., 1937, *J. Soc. Glass Tech.*, **21**, 442.

STOTT, V. H., 1937, *J. Soc. Glass Tech.*, **21**, 356.

STOTT, V. H., IRVINE, E., and TURNER, D., 1925, *Proc. Roy. Soc.* A, **108**, 154.

TAMMANN, G., 1933, *Der Glaszustand* (Leipzig : Teubner).

TAYLOR, N. W., 1941, *J. Appl. Phys.*, **12**, 753.

TAYLOR, N. W., and DORAN, R. F., 1941, *J. Amer. Ceram. Soc.*, **24**, 103.

TAYLOR, N. W., MCNAMARA, E. P., and SHERMAN, J., 1937, *J. Soc. Glass Tech.*, **21**, 61.
TAYLOR, N. W., and NOYES, Jr., B., 1944, *J. Amer. Ceram. Soc.*, **27**, 37.
TOOL, A. Q., 1942, *Bull. Amer. Ceram. Soc.*, **21**, 13 ; 1945, *Bur. Std. J. Res., Wash.*, **34**, 199 ; 1946, *Ibid.*, **37**, 73.
TOOL, A. Q., and EICHLIN, C. G., 1920, *J. Opt. Soc. Amer.*, **4**, 340.
TOOL, A. Q., and VALASEK, J., 1920, *Sci. Papers Bur. Std., Wash.*, **358**, 537.
TOOLEY, F. V., and TIEDE, R. L., 1944, *J. Amer. Ceram. Soc.*, **27**, 42.
TWYMAN, F., 1917, *J. Soc. Glass Tech.*, **1**, 61.
VOIGT, W., 1928, *Lehrbuch der Krystallphysik* (Leipzig : Teubner).
WARREN, B. E., 1933, *Z. Kristallogr.*, **86**, 349.
WARBURG, E., 1884, *Ann. Phys. Chem.*, **21**, 622.
WARD, A. G., 1937, *Trans. Faraday Soc.*, **33**, 88.
WEIDMANN, G., 1886, *Ann. Phys. Chem.*, **29**, 214.
WEYL, W. A., 1947, *Research*, **1**, 50.
WHITEHEAD, S., 1946, *Trans. Faraday Soc.*, **42**, 66.
WIECHERT, E., 1893, *Ann. Phys. Chem.*, **50**, 335.
WINTER, A., 1943, *J. Amer. Ceram. Soc.*, **26**, 189.
ZACHARIASEN, W. H., 1932, *J. Amer. Chem. Soc.*, **54**, 3841.
ZENER, C., 1948, *Elasticity and Anelasticity of Metals* (Chicago : University Press).

THEORY OF THE OXIDATION OF METALS

By N. CABRERA and N. F. MOTT
H. H. Wills Physical Laboratory, University of Bristol

CONTENTS

	PAGE
§ 1. Introduction	163
1.1. Formation of stable films at low temperatures	163
1.2. Region of intermediate temperatures	164
1.3. Parabolic law	164
§ 2. The parabolic law for thick films	166
§ 3. Theory of formation of thin films	173
§ 4. Formation of very thin films	178
4.1. Numerical values for an oxide which forms an excess semiconductor (aluminium)	180
4.2. Oxides which absorb oxygen; the case of copper	181
§ 5. Adhesion and crystal form of an oxide film	181

§ 1. INTRODUCTION

Recent theoretical and experimental work on the oxidation of metals has provided a general theoretical frame into which it may be possible to fit the complicated phenomena observed. While this scheme is by no means complete or proved at all points, it seems worth while to publish it in its present stage, in the hope that it may act as a guide to future experimental work.

In this Report the phenomena observed will be classified as follows:

1.1. *Formation of stable films at low temperatures.* Recent experimental work indicates that many metals, perhaps all which oxidize readily, show very similar behaviour when exposed to oxygen at a sufficiently low temperature. Oxidation is initially extremely rapid, but after a few minutes or hours drops to very low or negligible values, a stable film being formed 20–100 A. thick. Aluminium behaves like this at room temperature; copper, iron, barium and other metals do the same at the temperature of liquid air. An explanation of this behaviour was first given by Mott (1947 a), and depends on the hypothesis that a strong field is set up in the oxide film, due to a contact potential difference between metal and adsorbed oxygen, which enables the metal ions to move through it without much help from temperature; the theory gives a logarithmic growth law of the type

$$1/X = A - B \ln t,$$

X being the thickness at a time t. This mechanism is discussed in § 4 of this Report.

In many cases there is strong evidence that these films have a pseudomorphic form, are thus not in thermodynamic equilibrium, and are in fact highly compressed. The theoretical basis for understanding this phenomenon has been given by Frank and van der Merwe (1949 a, b) and van der Merwe (1949). According to these authors, the question whether the film will be pseudomorphic depends on the degree of fit or misfit between the lattice in the face of the metal crystal exposed and the spacing of the metal atoms in the oxide. They consider that a monolayer

of oxygen atoms will be formed very rapidly (say in 10^{-2} sec. at a pressure of 10^{-4} mm. Hg); for purposes of calculation they then discuss a metal covered by a monolayer of oxide. They then ask whether or not this monolayer of oxide will take up the lattice parameter of the metal underneath. The answer depends, of course, on the assumptions made about the forces between the oxide layer and the metallic substrate, and also on the elastic constants of the oxide layer itself; the more compressible the oxide, the more likely it will be to take up the distorted form. Making reasonable assumptions, Frank and van der Merwe find that, if the degree of misfit is less than about 15%, the film will, in its state of lowest energy, take up the lattice parameter of the substrate; if it is greater than 15%, it takes up very nearly its own unstrained lattice parameter.

In the latter case an oxide layer will be formed which, while it may have one crystal plane parallel to the surface layer of the crystal, will *not* have its crystal parameter distorted. In the former case, however, once a monomolecular layer of oxide is formed *all over* the surface, it must continue to grow with the same lattice parameter, even though, as soon as the layer thickens, the equilibrium becomes unstable. Provided that the surface is completely covered, the film can only assume its unstrained lattice parameter either by breaking away from the surface through plastic deformation or by recrystallization.

1.2. *Region of intermediate temperatures.* The discussion of the previous paragraph shows that a stable film will grow until it reaches a limiting thickness and will then stop, if the temperature is low enough for the following conditions to be satisfied: (*a*) metal ions cannot cross the film without the aid of a strong electric field, which only exists in *thin* films; (*b*) in the case of films compressed to fit the metal substrate the temperature at which crystallization occurs is not reached.

Recrystallization is a phenomenon which depends on the formation of one or more nuclei, and is thus a process likely to involve a long induction period. Figure 1, which gives results obtained at Bristol by Mrs. Garforth (1949) on the growth rate of oxide on an evaporated copper film determined with a quartz microbalance, shows a phenomenon probably to be explained by the recrystallization of a pseudomorphic film. It would appear from the curve that the recrystallized oxide can pass metal ions without the help of a strong field, so the curve will eventually go over into one of the forms described in paragraph 1.3 below; if the recrystallized material were still opaque, one would expect the logarithmic growth law of the type first described by Evans (1946) for films a few 100 A. thick (not to be confused with the mechanism described under 1.1 above).

For metals for which the original oxide film is not fitted to the metal substrate (aluminium), and perhaps for recrystallized films too, another intermediate region can be recognized, that in which the temperature is high enough for ions to diffuse without the help of a strong field, but in which the thickness is not great enough, in the times used in the experiment, for the parabolic law ($X^2 = 2At$) to be valid (see 1.3 below). In this region it is possible to find theoretical justifications for various laws, a parabolic law with a different constant A, a cubic law $X^3 = 3At$ (cf. the work of Campbell and Thomas (1947) on copper), a roughly linear law $X = At$ (cf. Gulbransen and Wysong (1947) for Al). A brief discussion of these laws will be given in § 3.

1.3. *Parabolic law.* For sufficiently high temperatures, *and* sufficiently thick films, the oxidation should conform to the parabolic law. The derivation of this

law ($X^2 = 2At$) depends on the following assumptions. Either metal or oxygen is soluble in the oxide; local thermodynamic equilibrium exists at the metal–oxide interface and at the oxide–air interface; the concentrations of metal (or oxygen) at the two faces are therefore different; metal or oxygen thus diffuses through the oxide layer under a concentration gradient which is proportional to $1/X$; the rate of growth dX/dt is thus proportional to $1/X$, and integration gives the parabolic law.

Oxides such as those of zinc and aluminium do not dissolve oxygen; they can, however, dissolve metal (thereby becoming excess semiconductors). For such oxides, then, one can assume a vanishing concentration of metal at the oxygen–air interface even for low pressures of oxygen; the rate of oxidation is thus independent of oxygen pressure. Oxides such as those of copper and iron dissolve oxygen

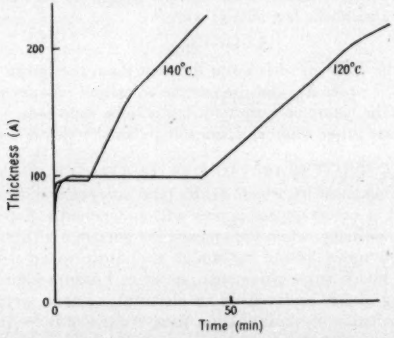


Figure 1. Rate of growth of oxide layer on evaporated copper film at 120° c. and 140° c., at a pressure of oxygen of 1–3 mm. Hg. The weights of oxygen taken up were determined by the microbalance, and converted to thicknesses using the known density of the oxide.

(thus becoming defect semiconductors); the concentration gradient of oxygen thus varies with oxygen pressure, as does also the rate of oxidation. Nevertheless, it is not the oxygen which diffuses, but the metal, as has been proved by the use of radioactive tracers in the case of copper (Bardeen, Brattain and Shockley 1946). This is because the oxygen is taken up in such a way as to form vacant cation sites, which diffuse away from the oxide–air interface.

In trying to estimate the value of the constant A, one must remember that the dissolved metal atoms are almost completely dissociated, e.g. into interstitial ions and electrons. The value of A will depend on whether the pure (stoichiometric) oxide is an insulator at the temperature considered. If this is the case, we shall show in the next section that

$$A = 2\Omega D_i n, \qquad \ldots\ldots(1)$$

where Ω is the volume of oxide per metal ion, D_i the diffusion coefficient for an interstitial ion and n the concentration of dissolved atoms (ions and electrons) at the interface.

The derivation of this formula is valid only if the film is so thick that the concentrations of ions and electrons are equal throughout most of its thickness.

Actually at each boundary there will be layers where they are unequal, and where, in consequence, a space charge is set up, giving rise to a double layer. The thickness of this layer is of order $X_0 = \sqrt{(\kappa \boldsymbol{k} T/8\pi n \boldsymbol{e}^2)}$, where κ is the dielectric constant. This constant is of course very sensitive to temperature. As will be shown in §3, in practice, in experiments lasting a few hours, if the thickness X has grown to 10^{-4} cm., then $X \gg X_0$, and the condition for the parabolic law is satisfied.

For films of thickness less than X_0 the space charge set up in the material, if the concentrations of ions and electrons are unequal, has very little effect. One can thus discuss the motion of ions and electrons separately. The much more mobile electrons will probably pass freely through the film and set up a constant potential difference V between the metal and the adsorbed oxygen layer. The field in the oxide film is thus V/X. If, then, n_i is the concentration of metal *ions* in solution and v_i their mobility, the current is $n_i v_i V/X$ ions/cm^{2}sec., and it is easily seen that a parabolic law follows with

$$A = \Omega v_i V n_i. \qquad \ldots\ldots(2)$$

This value of the constant A is quite different from that given by (1), and for thicknesses of the order X_0, and thus in the transition regions between the two laws nothing in the nature of a parabolic law is to be expected.

Case 1.1 above arises when n_i is vanishingly small; this is discussed in §4.

§2. THE PARABOLIC LAW FOR THICK FILMS

2.1. The mechanism by which oxides (and sulphides and halides) can take up excess metal or excess oxygen is now well understood. Excess metal can be taken either interstitially, when the metallic ion goes into an interstitial position and the electron moves round it through the lattice in an "orbit" probably extending over many lattice parameters, or as an F-centre (site from which the anion is missing and is replaced by an electron). Excess oxygen is taken up through the formation of vacant cation sites, a positive hole (missing electron) being located on an adjacent ion.

An oxide containing excess metal (or oxygen) quenched to a temperature at which the interstitial ions or vacant sites are not mobile behaves of course as an electronic semiconductor*; at low temperatures the electrons are bound to their interstitial ions, but as the temperature is raised an increasing proportion becomes free. When, however, an oxide is in thermodynamic equilibrium with its metal, at a temperature at which the interstitial ions are mobile, the dissolved atoms will nearly all be dissociated into interstitial ions and electrons. It will be assumed throughout this Report that dissociation is complete.

2.2. We have then to discuss the contact between a metal and an oxide capable of accepting excess metal, and to determine the concentration n_i of interstitial ions and n_e of electrons in the oxide at a distance x from the interface, when the whole system is in thermodynamical equilibrium. If x is large enough, n_i and n_e become equal, but this is not so for small x, so that a space charge is set up at the boundary. At the boundary we envisage a situation such as that shown in Figure 2; the process of solution of an ion in the oxide is typically the removal of the ion at P from its position in the surface layer of the metal into an interstitial position of the oxide. If the energy required to do this is denoted by W_i, then W_i may be described as the heat of solution of a metallic ion.

* For a recent eview of the theory of semiconductors see Mott (1949 a).

We may also introduce the energy ϕ required to remove an electron from the metal into the conduction band of the oxide; for oxide grown chemically on the metal we may expect this to have a characteristic value, adsorbed gas layers being absent from the interface.

The usual diagram for an insulator in contact with a metal is shown in Figure 3; ϕ will in general be less than the work function of the metal against vacuum.

The quantity $\phi + W_i - \epsilon$ is the heat of solution of the metal atom in the oxide; here ϵ is the energy with which an electron is bound to the interstitial ion in the oxide. Thus, while ϕ and W_i may individually depend on which crystal face of the metal is exposed, $\phi + W_i$ will not.

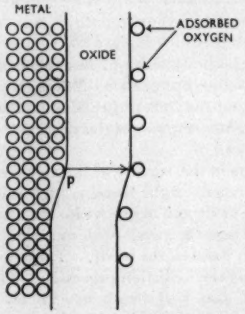


Figure 2. Showing the mechanism by which ions leave a metal and pass through oxide layer (Mott 1947 a).

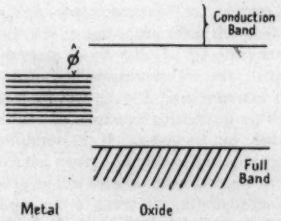


Figure 3. Metal in contact with insulating oxide.

It is easily shown that, if ϵ is less than both ϕ and W_i by at least several multiples of kT, the dissociation is almost complete; this will be assumed in what follows.

Immediately at the interface the concentration $n_i(0)$ of interstitial ions is given by

$$n_i(0) = N_i \exp(-W_i/kT), \qquad \ldots\ldots(3)$$

where N_i is the number of interstitial positions per unit volume in the oxide. Similarly the number $n_e(0)$ of electrons is

$$n_e(0) = N_e \exp(-\phi/kT), \qquad \ldots\ldots(4)$$

where $N_e = 2(2\pi mkT/h^2)^{3/2}$. Also if $n_i(x)$, $n_e(x)$ are the numbers at any distance x from the interface, the product $n_i(x)n_e(x)$ must be constant; thus at large distances $n_i(x)$ and $n_e(x)$ are equal, say, to n, where

$$n = \sqrt{(N_i N_e)} \exp\{-\tfrac{1}{2}(W_i + \phi)/kT\}. \qquad \ldots\ldots(5)$$

At intermediate distances these quantities can be deduced from Boltzmann's law:

$$n_i(x) = n \exp(-eV/kT);$$

$$n_e(x) = n \exp(eV/kT),$$

where V is the electrostatic potential, and from Poisson's equation

$$\frac{d^2V}{dx^2} = \frac{4\pi e}{\kappa}\{n_i(x) - n_e(x)\}.$$

It is assumed here that ions carry unit charge.

Substituting for n_i, n_e, this gives

$$\frac{d^2V}{dx^2} = \frac{8\pi n}{\kappa} e \sinh\left(\frac{eV}{kT}\right). \qquad \ldots\ldots(6)$$

An exact solution satisfying the boundary conditions (3), (4) can easily be obtained (cf. Mott 1947 b), but is not necessary for our purpose. We need only examine the form of equation (6) when x is large and V consequently small; it then becomes $d^2V/dx^2 = V/X_0^2$, where

$$X_0 = \sqrt{\{\kappa kT/8\pi ne^2\}}. \qquad \ldots\ldots(7)$$

The solution is $V = \text{const.}\exp(-x/X_0)$. This shows that our treatment of the problem will differ according as the thickness X of our growing film is or is not greater than X_0. If $X \gg X_0$ we may, throughout the bulk of the film, treat n_e, n_i as equal; this assumption will be made throughout the rest of this section. The other extreme case, $X \ll X_0$, will be treated in §4.

Other interesting examples of double layers in the vicinity of the metal–oxide interface can be given. If the solubility of oxygen in the metal is high and the energy required to take an oxygen ion from the oxide and to put it into the metal is not too big, vacant anion sites will be produced near the metal–oxide interface, with the corresponding electrons in the conduction band of the oxide. These vacant anion sites will diffuse through the oxide during the oxidation processes. This is probably the mechanism by which Ag and also Cu absorb oxygen at high temperatures without the growth of a thick layer of oxide, as long as the saturation concentration of dissolved oxygen in the metal has not been reached.

2.3. The case of an oxide in equilibrium with the oxygen gas, and the formation of a double layer near the oxide–air interface, can be treated in a similar way. Let us consider the case of an oxide (Cu_2O) absorbing oxygen through the formation of vacant cation sites (concentration n_i) and positive holes (concentration n_e) in the full band of the oxide. Far from the oxide–air interface both concentrations are equal to n, given by (see Mott and Gurney 1948, p. 260)

$$n = \sqrt{(N_i N_c)}\left(\frac{n_g}{N_g}\right)^{1/2s}\exp(-E/2skT), \qquad \ldots\ldots(8)$$

where N_i and N_c are as in §2.2, and

$$N_g = (2\pi MkT/h^2)^{3/2};$$

M is the mass of an oxygen molecule, n_g the number of oxygen molecules per cm^3 in the air and E the energy required for the absorption of one of them and the formation of s vacant cation sites and s positive holes ($s = 4$ in the case of Cu_2O). The concentration n will therefore be proportional to $p^{\frac{1}{4}}$, where p is the pressure.

The calculation of the concentrations $n_i(0)$ and $n_e(0)$ near the oxide–air interface requires more careful consideration. Figure 4 represents the electronic levels of the oxide and those of the adsorbed oxygen layer (we assume for simplicity that the levels of the adsorbed layer correspond to the same energy). As we shall see in §4, the adsorbed levels may be below the top of the full band (ψ might be negative), particularly in the case of Cu_2O.

Let N_0 and N be the number of empty and full levels in the adsorbed layer per cm^2. Then it is easy to see that $n_e(0)$ is given by

$$n_e(0) = N_e(N_0/N)\exp(-\psi/\boldsymbol{k}T),$$

where the factor N_0/N is due to the different ways in which the N occupied levels can be distributed among the total number N_0+N of levels. On the other hand, the vacant cation sites will be produced *only* in the neighbourhood of the N adsorbed oxygen ions, requiring an energy W_i, according to Figure 5. Therefore $n_i(0)$ will be given by $n_i(0) = N_i(Na^2)\exp(-W_i/\boldsymbol{k}T)$, where Na^2 represents the proportion of the oxide–air interface occupied by oxygen ions.

The presence of the factor (a^2N) in the formula above can be seen by an application of the principle of detailed balancing. The number of vacant cation sites created per second will be proportional to (a^2N); the adsorbed oxygen ion is then neutralized and another one is produced somewhere else in order to keep N constant. On the other hand the number of vacant cation sites disappearing per second will be proportional to $n_i(0)/N_i$ and independent of (a^2N), because this process *does not* require the vicinity of an adsorbed oxygen ion.

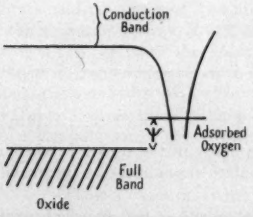


Figure 4. Insulating oxide in contact with oxygen.

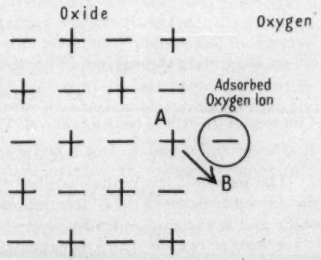


Figure 5. Showing the process by which a vacant cation site is created in the vicinity of an adsorbed oxygen ion.

As $n_e(0)n_i(0) = n^2$, we deduce that the proportion of the oxide–air interface covered with adsorbed neutral oxygen is

$$N_0a^2 = \left(\frac{n_g}{N_g}\right)^{1/s}\exp\left\{-\left(\frac{E}{S}-\psi-W_i\right)\Big/\boldsymbol{k}T\right\},$$

obviously a function of pressure. On the other hand the number N of adsorbed oxygen ions will be equal to the total space charge near the oxide–air interface, and, therefore, also a function of pressure.

Throughout this discussion it has been assumed that the oxide of stoichiometric composition is an insulator i.e. it has no intrinsic electronic or ionic conductivity, or these conductivities are small compared with that due to the dissolved atoms.

2.4. *Calculation of rate of growth of oxide film.* In the oxide film, we denote by D_e, v_e the diffusion coefficient and mobility of an electron and D_i, v_i the same

quantities for an interstitial ion. Then if F is the field in the film, the current j_e carried by electrons is (in units of e)

$$j_e = -D_e \frac{\partial n_e}{\partial x} + F n_e v_e,$$

and the current j_i carried by the ions

$$j_i = -D_i \frac{\partial n_i}{\partial x} - F n_i v_i.$$

In a steady state these are equal and opposite; thus, putting both equal to j (number of atoms crossing unit area per unit time), we see on eliminating F between the two equations,

$$j\left\{\frac{1}{n_e v} + \frac{1}{n_i v_i}\right\} = -\frac{kT}{e}\frac{\partial}{\partial x}\ln(n_e n_i). \qquad \ldots\ldots(9)$$

In deriving this equation use has been made of the Einstein equation $D/v = kT/e$.

We may certainly assume that $v_e \gg v_i$. In the case of an oxide which is an insulator in the absence of dissolved metal or oxygen, and for which $X \gg X_0$, we may further assume that $n_i(x) = n_e(x) = n(x)$, say, throughout the film, except for the boundary zone, which we neglect. The equation (9) thus becomes

$$j = -2D_i \frac{\partial n}{\partial x},$$

or, on integrating throughout a film of thickness X,

$$jX = 2D_i[n(0) - n(X)].$$

The rate of growth is thus $dX/dt = A/X$, whence $X^2 = 2At$, where

$$A = 2D_i\Omega[n(0) - n(X)]. \qquad \ldots\ldots(10)$$

The parabolic law is thus satisfied. The quantity in the square brackets is the difference between the concentrations of dissolved atoms at the surface metal–oxide and at the surface oxide–oxygen.

A further case in which equation (9) can be solved is that in which oxide (or halide) is a good ionic conductor in the pure state, and the electronic conductivity, due to the addition of metal, is small in comparison. This may be the case for the halides. We may then set $n_e v_e \ll n_i v_i$ and take n_i as constant throughout the film. The equation then becomes

$$j = -v_e \frac{kT}{e}\frac{\partial n_e}{\partial x},$$

so the parabolic law is satisfied with

$$A = D_e\Omega[n_e(0) - n_e(X)].$$

Returning to (10), the usual assumption is that, for oxides which form excess semiconductors (ZnO, Al_2O_3), the concentration $n_e(X)$ of dissolved metal at the oxide–air interface is zero (except for very small pressures of oxygen), so that the oxidation constant becomes

$$A = 2D_i\Omega n(0). \qquad \ldots\ldots(11)$$

Note that here D_i is the diffusion coefficient of the interstitial ions and $n(0)$ the concentration of metal *atoms* (dissociated or otherwise) in the oxide in equilibrium with metal. For oxides such as Cu_2O on the other hand we assume that $n(0)$,

the concentration of vacant sites in equilibrium with the metal, is small compared with that, $n(X)$, in equilibrium with the air; thus, approximately,

$$A = 2D_i \Omega n(X), \qquad \ldots\ldots(12)$$

where $n(X)$ is the concentration of vacant cation sites in the oxide in equilibrium with oxygen, and D_i the diffusion coefficient for one of these sites. Note that in the former case the oxidation rate is almost independent of oxygen pressure p, unless this is very low; in the latter case, according to (8), it should vary as $p^{\frac{1}{2}}$.

2.5. *Experimental results at high temperatures.* Putting in formulae (11) or (12) the usual expression $D_i \sim a^2 \nu \exp(-U/\boldsymbol{k}T)$ for the diffusion coefficient of defects in the oxide, a being the interatomic distance, $\nu (\sim 10^{12}\,\mathrm{sec}^{-1})$ the atomic frequency of vibration and U the activation energy for movement between two equilibrium positions of the defect, and using formulae (5) or (8) for n, we get, for the constant A in the parabolic law, an expression of the form

$$A = A_0 \exp(-B/\boldsymbol{k}T), \qquad \ldots\ldots(13)$$

where

$$A_0 = 2a^2 \nu \Omega \sqrt{(N_i N_e)}, \qquad B = \tfrac{1}{2}(W_i + \phi) + U \text{ or } E/2s + U.$$

The factor $(n_s/N_s)^{1/2s}$ appearing in (8) is of the order unity and can be disregarded. Taking $a \sim 3 \times 10^{-8}$ cm., $\nu \sim 10^{12}\,\mathrm{sec}^{-1}$, $\Omega \sim a^3$, $N_i \sim 10^{22}\,\mathrm{cm}^{-3}$, $N_e \sim 10^{18}\,\mathrm{cm}^{-3}$, we obtain $A_0 \sim 10^{-5}\,\mathrm{cm}^2/\mathrm{sec}$. In this evaluation we have disregarded the change of vibrational frequencies of the solid due to the presence of defects and the change of the activation energy B with temperature. This may, however, introduce a factor which can be of the order of a few powers of 10 (see Mott and Gurney 1948, p. 29), and which seems to be larger for the vacancy type of defect than for interstitial ions.

Table 1 gives the constants A_0 (in cm^2/sec.) and B (in ev.), determined experimentally during the oxidation of several metals in the range of temperature indicated. In some of these cases the oxide layer is complex.

Table 1. Values for the Constants A_0 and B in Formula (13) and governing Rate of Oxidation

	Temperature range	A_0 (cm^2/sec.)	B (ev.)
Fe	700° to 900° c.	1·0	1·6
Cu	700° to 1000° c.	0·2	1·5
Ni	800° to 1000° c.	0·1	2
Pb	470° to 626° c.	0·2	1·4
Zn	600° to 700° c.	0·002	1·5

For Fe there are three layers (FeO, Fe_3O_4, Fe_2O_3, going from metal to air). The values given in Table 1 correspond to the thickness of FeO (Bénard and Coquelle 1946), which in the range of temperatures considered forms 90% of the total thickness. It is believed that the diffusing elements are mostly vacant cation sites formed at the FeO–Fe_3O_4 interface. For Cu, at pressures below 100 mm. Hg at 1,000° c., there is only Cu_2O and no CuO. Assuming that the diffusing elements are also vacant cation sites produced at the Cu_2O–air interface, one expects from formula (8) that A_0 should be proportional to $p^{\frac{1}{2}}$; this was proved to be the case by Wagner and Grünewald (1938). The values given in Table 1 are extrapolated to $p = 1$ atm. For Ni, vacant cation sites are also responsible for the diffusion of the metal, and, therefore, A_0 is also a function of the pressure, as was proved by Wagner and Grünewald. The values quoted in

Table 1 are deduced from data given by Evans (1948, p. 144). The values for Pb and Zn were observed by Krupkowski and Balicki* (1937), who studied the formation of solid oxide layers on the surface of liquid metals. It is well known that the diffusing elements in ZnO are interstitial ions formed at the metal–oxide interface; therefore one expects the constant A_0 to be independent of the pressure, as was shown by Wagner and Grünewald (1938).

Bénard and Talbot (1948) have studied also the oxidation of single crystals of copper at 900° c. Figure 6 represents the increase of weight as a function of time on different faces of a copper crystal. These differences were also reported earlier by Gwathmey and Benton (1940) on copper at 1,000° c. It is known from the work of Wagner and Grünewald (1938) that in the case of copper the con-

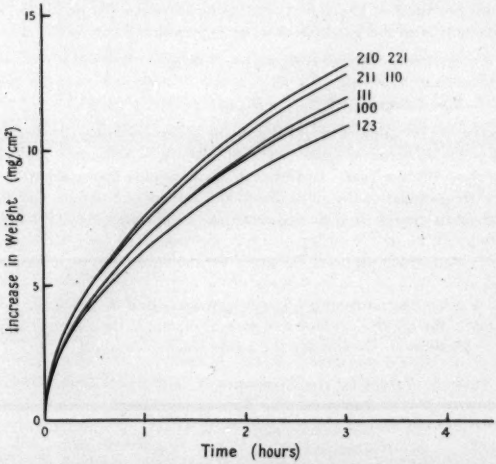


Figure 6. Oxidation of different faces of copper at 900° c. (Bénard and Talbot 1948).

centration of vacant cation sites in the oxide near the metal–oxide interface $n(0)$ is not quite negligible in comparison with their concentration near the oxide–air interface $n(X)$. Therefore formula (10) has to be used. On the other hand, the diffusion coefficient D_1 cannot be a function of direction in a cubic crystal such as Cu_2O; therefore, Bénard and Talbot's results show that $n(0)$ or $n(X)$, or perhaps both, are not equal to the equilibrium concentrations, which of course would not depend on the crystal surfaces exposed. As the differences depend on the orientation of copper surface we expect that $n(0)$ will not be equal to the equilibrium concentration, but will be higher for the faces for which the rate of oxidation is lower.

One of us (Cabrera 1949 b) has advanced the following hypothesis to explain this. As long as there are enough positions such as P (Figure 2) on the metal

* These authors observed also the oxidation of liquid Cu (1,100° to 1,200° c.) and Ag (1,000° to 1,035° c.), for which they obtained an initial linear increase of weight without the formation of a thick oxide layer, due to the absorption of oxygen in the liquid metal.

surface, the concentration of vacant cation sites near the metal will be maintained equal to the equilibrium concentration, in spite of the constant arrival of new vacant cation sites; but when the evaporation into the oxide of a new atomic layer of the metal has to be started, a larger activation energy will be required for the formation of a "hole" on the metal surface, such as is illustrated in Figure 7, and, therefore, the concentration of vacant cation sites in the oxide will grow.

This point of view assumes that the metal surface has a practically perfect structure, and one can deduce that the order of increasing rate of oxidation should be the same as the order of decreasing density of metal atoms on the surface: that is to say (111), (100), (110), (311), (331), which is not in agreement with the experimental results represented in Figure 6. Actually the metal surface is not

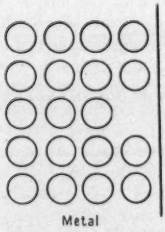


Figure 7. First step in the evaporation of a new layer of metal into the oxide.

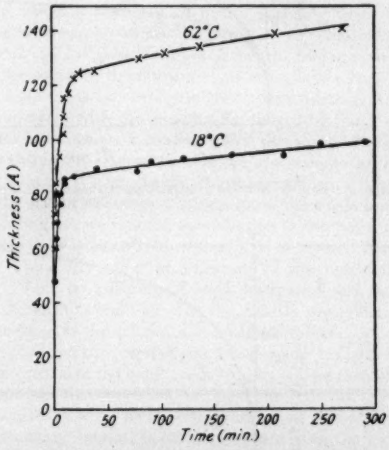


Figure 8. Growth curves for Cu_2O on Cu determined by Miley and Evans by an electrometric method.

even microscopically perfect; according to an idea put forward by Frank and others (Burton, Cabrera and Frank 1949, Frank 1949) a certain type of "dislocation" in the body of the metal produces positions such as P (Figure 2) on the metal surface, which are *not* destroyed when the entire atomic layer is evaporated; therefore the rate of oxidation on different crystal surfaces will depend on the concentration of these dislocations per unit area of each crystal surface. Just how this will depend on the crystal face is not at present clear.

§3. THEORY OF FORMATION OF THIN FILMS

In this section we outline the theory of the rate of growth of thin films. By thin we mean of thickness X small compared with X_0 defined by (7), so that the concentrations of positive ions and electrons diffusing across the layer can become unequal without any important space charge being set up. In practice most of the films that we shall discuss are of thickness 100 A. or less.

No attempt will be made here to survey all the available experimental material. A survey of all but the most recent material is given in the book by Evans (1948); Figure 8 shows results, taken from p. 65 of this book, on the oxidation of copper, measured by an electrometric method. As an example of more recent work Figure 9 represents the rate of oxidation of thin films of aluminium at 10° C.

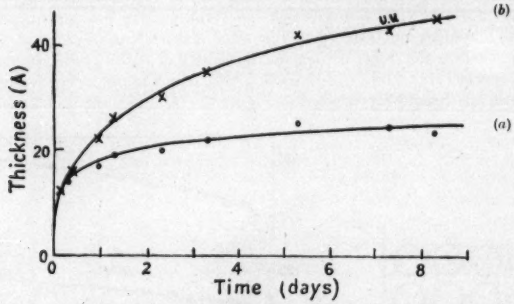


Figure 9. Rate of oxidation of aluminium (Cabrera *et al.* 1947) (*a*) in the dark, (*b*) under ultra-violet illumination.

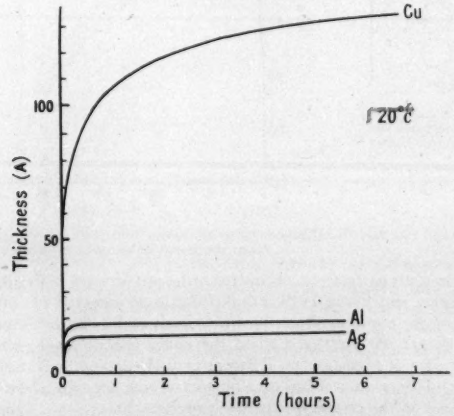


Figure 10. Oxidation at room temperature and pressures of oxygen between 10^{-4} and 10^{-2} mm. Hg (Garforth 1949).

These results were obtained by Cabrera, Terrien and Hamon (1947) from the increase in the transparency of the film during oxidation, a method first used by Steinheil (1934). The upper part of the curve represents the influence on oxidation of ultra-violet light, which will be discussed later. Figure 10 shows the oxidation of copper, aluminium and silver at room temperature and oxygen

pressure of 10^{-4} to 10^{-2} mm. Hg measured recently in this laboratory by Mrs. Garforth (1949) using a quartz microbalance.

The general result of all this work is that all metals investigated show in principle a similar behaviour. If the temperature is low enough, they show, when exposed to oxygen, an initial very rapid growth, followed by a remarkable slowing down, and for some critical thickness X_L of order 100 A. or less growth stops or nearly stops. This behaviour is rather insensitive to the pressure of oxygen.

In this section and the next, then, we shall discuss the mechanism by which these very *thin* films of oxide grow. Consideration of the mechanism by which the first monolayer is formed is deferred until §5, as is also the related question of whether or not the film is pseudomorphic (strained). Here we suppose that a thin layer of oxide exists in the metal and is growing; we require to know how fast it grows. As in §2, the oxide is supposed to be an insulator when of stoichiometric composition but to be capable of dissolving metal ions; in contradistinction to the case considered in §2, it is here supposed that the film is so thin that the effect of any space charge set up by the dissolved ions is negligible, so that the movements of ions and of electrons can be considered independently. In other words, the film is thin compared with the quantity X_0 defined by (7).

It will be worth while to consider again the condition that this should be the case. Suppose that n_i is the number of interstitial ions per unit volume in the oxide when in equilibrium with metal, and let q be the charge on each ion. Then Poisson's equation gives for the potential energy V of an ion in the oxide layer $d^2V/dx^2 = 4\pi n_i q^2/\kappa$. On integrating we see that the contribution to the potential due to the space charge alone is $V = 2\pi n_i q^2 x^2/\kappa$, where x is measured from the mid-point of the film. Thus for a film of thickness X the maximum variation of V is $\frac{1}{2}\pi n_i q^2 X^2/\kappa$. This is negligible if small compared with $\boldsymbol{kT}$. Thus, omitting numerical factors, the condition that the field is negligible is $X \ll \sqrt{(\kappa \boldsymbol{k} T/n_i q^2)}$. This is the same as the relation (7) already obtained.

We consider then a film of oxide on the metal exposed to oxygen. A layer of oxygen will be adsorbed to the surface of the oxide; this oxygen will be assumed to be atomic. We assume further that electrons can pass through the oxide layer from the metal to the oxygen by some mechanism (thermionic emission or tunnel effect), and that the electronic motion is rapid compared with the ionic motion. Some of the adsorbed oxygen atoms will then be converted into ions O^-, setting up a field across the oxide layer, until a state of quasi-eqilibrium is set up between the metal and the adsorbed oxygen, in which, in a time short compared with that in which the metal ions diffuse, as many electrons pass in one direction as the other. The electrostatic potential V set up across the layer will clearly be independent of the thickness, so that the field F is given by $F = V/X$.

The electronic levels in the metal, the oxide and in the adsorbed oxygen layer are shown in Figure 11. Figure 11(*a*) shows the state of affairs before any electrons have passed through the film; in the oxygen atoms there are energy levels *below* the surface of the Fermi distribution in the metal*; the quantity eV already introduced is the amount by which they are lowered. Electrons will pass through the film until the quasi-steady state in Figure 11(*b*) is reached, when as many electrons pass through the film in one direction as in the other. Since V in practice is of the order of one or two volts, its variation with temperature can be ignored.

* These energy levels are probably the surface states described by Bardeen (1947) in his work on rectifiers, cf. Mott (1949 b).

A very rough numerical estimate of V can be given. If E is the electron affinity of O and W_{bind} the adsorption energy of an oxygen ion O^- on the surface of oxide, then $eV = E + W_{\text{bind}} - \phi_0$, where ϕ_0 is the work function of the metal against vacuum. We do not know very much about W_{bind}; if one considers the oxides as ionic solids, the calculations of Lennard-Jones and Dent (1928) suggest that $W_{\text{bind}} \sim 0{\cdot}1e^2/2r$, where r is the radius of the ion O^-. This gives $W_{\text{bind}} \sim 1$ ev. if $r \sim 1$ A. This value is probably too low by a factor 2 or 3, because of the covalent forces existing in the oxides. Taking (Bates and Massey 1943) $E = 2{\cdot}2$ ev. and $\phi_0 = 4{\cdot}6$ ev. for copper, or $4{\cdot}3$ ev. for aluminium, we get values for V of the order of 1 volt.

The growth of the film is due to the strong field set up in this way, which pulls the interstitial metal ions through it. These fields may be very strong; thus for a film 50 A. thick they may be of the order 10^7 v/cm. For these fields the diffusion velocity of a positive interstitial ion is no longer proportional to the field. The condition that the diffusion velocity shall be proportional to the field F is $qaF \ll kT$, where q is the charge on the ion, a the distance between interstitial

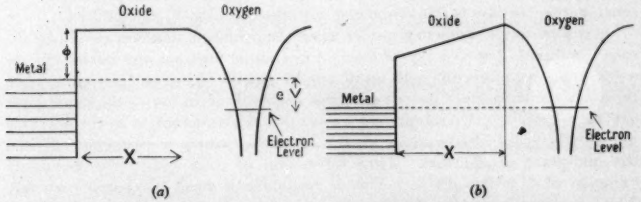


Figure 11. Electronic levels in the metal, oxide and adsorbed oxygen: (*a*) before electrons have passed through the oxide, (*b*) when equilibrium is set up.

positions. This will be the case for films of thickness X large compared with X_1, where $X_1 = qaV/kT$. X_1 is of order 50–200 atomic layers at room temperatures.

We treat first the case when $X \gg X_1$ (though less than X_0 defined by (7)). The number of ions crossing unit area per unit time is

$$n_i v_i F = n_i v_i V/X,$$

where v_i is the mobility of an ion. The rate of growth is thus $dX/dt = A/X$, where

$$A = n_i v_i V\Omega = n_i D_i \Omega(eV/kT). \qquad \ldots\ldots(14)$$

The film thus grows according to the parabolic law. Actually Gulbransen and Wysong (1947) have observed for aluminium a parabolic law between 350° and 450° C., which corresponds very probably to the case considered above. It is, of course, always assumed that electrons can pass freely through film, by either thermionic emission or tunnel effect. If ϕ (Figure 3) is too great for thermionic emission and the film too thick for tunnel effect (*c.* 30 A., cf. Mott 1940), these formulae may be expected to break down and the film to stop growing. Whether this corresponds to any case observed in practice is not known.

It is to be emphasized that the constant A given by (14) is quite different from that derived for $X \gg X_0$ (formula (11)). It depends on n_i, and thus on the exponential factor $\exp(-W_i/kT)$ instead of on $\sqrt{(n_e n_i)}$ and thus on $\exp\{-\frac{1}{2}(W_i + \phi)/kT\}$.

If, for instance, W_1 is larger than ϕ, then in the transition region where $X \sim X_0$ the value of A in the parabolic equation increases. It is likely that an increase of this type may account for the apparent linear law observed by Gulbransen and Wysong (1947) for aluminium in the region around 500° c.

It is easy to make a rough estimate of whether the thickness X of an oxide film is greater or less than the thickness X_0 above which the densities of electrons and positive ions can be taken as equal. Let us suppose a film to be growing according to a parabolic law

$$\tfrac{1}{2}X^2 = a^2\nu t \exp\{-(W+U)/\boldsymbol{k}T\}, \qquad \ldots\ldots(15)$$

where W is the heat of solution of an atom or ion and U the activation energy for motion. We want to know whether X^2 is greater or less than X_0^2 where

$$X_0^2 \sim (a^3\boldsymbol{k}T/\boldsymbol{e}^2)\exp(W/\boldsymbol{k}T),$$

and thus whether

$$\exp(-W/\boldsymbol{k}T) \lesseqgtr a^3\boldsymbol{k}T/\boldsymbol{e}^2X^2.$$

Substituting for $\exp(-W/\boldsymbol{k}T)$ from (15), the condition becomes

$$(X^2/2a^2\nu t)^\gamma \lesseqgtr a^3\boldsymbol{k}T/\boldsymbol{e}^2X^2, \qquad \gamma = W/(W+U),$$

and thus

$$\frac{X}{a} \gtreqless \left(\frac{a\boldsymbol{k}T}{\boldsymbol{e}^2}\right)^{1/2(\gamma+1)} (\nu t)^{\gamma/2(\gamma+1)}.$$

γ is probably about 2/3; the factor $(\boldsymbol{k}Ta/\boldsymbol{e}^2)^{1/2(\gamma+1)}$ is then about 4, so that $X/a \gtrsim 4(\nu t)^{1/5}$.

For experiments lasting a few hours ($t = 10^4$ sec.), and with $\nu = 10^{12}$ sec^{-1}, we see that the critical thickness X is about $6 \times 10^3 a$, or about 2×10^{-4} cm. If in a few hours the film has grown to this thickness the mechanism is that of § 2; if not, it is the mechanism of § 3.

Up to this point the theory has been developed for metals whose oxides form excess semiconductors, e.g. Zn, A. For these metals n_1 depends on the metal–oxide equilibrium, and is, of course, independent of thickness. For the oxides of copper, iron, and so on, which take up excess oxygen, the position is rather different. n_1 will then refer to the number of places where a cation is missing. These can be formed wherever an oxide ion is adsorbed to the surface by the process shown in Figure 5; W_1 now represents the energy required to move the positive ion from A to B. It is clear, then, since $n_1 = (N/a)\exp(-W_1/\boldsymbol{k}T)$, that the concentration n_1 is proportional to the number N of such negative ions per unit area.

On the other hand N is related to the field F by the Coulomb formula

$$N = \frac{\kappa F}{4\pi\boldsymbol{e}} = \frac{\kappa V}{4\pi\boldsymbol{e}}\frac{1}{X},$$

showing that N is proportional to X^{-1}; therefore n_1 in formula (14) is now proportional to X^{-1}. This leads to an oxidation law of the type

$$X^3 = 3At, \qquad \ldots\ldots(16)$$

where the constant A is proportional to $\exp\{-(W_1+U)/\boldsymbol{k}T\}$, W_1 being the energy to form a vacant cation site and U the activation energy for its

diffusion in the oxide. Cabrera (1949a) has used this formula to explain the cubic law observed by Campbell and Thomas (1947) on copper between 100° and 250° c. Their results can be expressed by a formula such as (16), where the constant A is given by

$$A \sim 10^{-7} \exp(-B/\boldsymbol{k}T)\,\mathrm{cm}^3/\mathrm{sec.}, \qquad B = W_1 + U \sim 1{\cdot}1\,\mathrm{ev}.$$

As noticed before (§2), the differences in the rate of oxidation for different crystal faces, for films of thickness X great compared with X_0 (*c.* 10^{-4} cm.), can be explained only on the basis of a surface nucleation process occurring at the metal–oxide interface, because $W_1 + \phi$, the heat of solution of metal, cannot depend on the crystal face. When $X \ll X_0$, the same surface nucleation process will play a rôle, but we expect also to have differences in the rate of oxidation due to the fact that W_1 and ϕ (or V) may well depend on the crystal face. Gwathmey and Benton (1942) have observed differences in the rate of oxidation on a spherical single crystal of copper at 200° c. The metallic crystal faces, ordered according to a decreasing rate of oxidation, are (100), (210), (111), (110), (311). The fact that this order is different from that reported by Bénard and Talbot (1948) on copper at 900° c. suggests that the rate-determining factor at low temperatures is the difference in W_1.

§4. FORMATION OF VERY THIN FILMS

For very thin films the field is so strong that the velocity of drift of the ions is no longer proportional to it. In this case the motion can be treated as follows (Mott 1947a). Suppose that an ion has to go over a potential barrier U in order to move from one interstitial site to the next. In the absence of a field the chance per unit time that an ion will do this is $\nu \exp(-U/\boldsymbol{k}T)$ with $\nu \sim 10^{12}\,\mathrm{sec}^{-1}$. The field will, however, lower the barrier by $\frac{1}{2}\,qaF$ for motion in the direction of the field, increasing the probability of movement to $\nu \exp\{-(U - \frac{1}{2}qaF)/\boldsymbol{k}T\}$. In the opposite direction the chance of movement is decreased by the same factor. Thus the velocity u of drift becomes

$$u = \nu a \exp(-U/\boldsymbol{k}T)\{\exp(\tfrac{1}{2}qaF/\boldsymbol{k}T) - \exp(-\tfrac{1}{2}qaF/\boldsymbol{k}T)\}.$$

For small values of F this reduces to

$$u \sim (\nu a^2 q/\boldsymbol{k}T)F \exp(-U/\boldsymbol{k}T),$$

which is proportional to the field. For large values, on the other hand,

$$u \sim \nu a \exp(-U/\boldsymbol{k}T) \exp(\tfrac{1}{2}qaF/\boldsymbol{k}T),$$

giving an exponential dependence.

It will be seen that, when the field is strong, the motion of the ions is overwhelmingly in one direction; there is, therefore, no question of any local equilibrium between metal and oxide, since equilibrium is only set up when there is a continual exchange of ions. Thus every ion which escapes from the metal is pulled right across the film, and none recombine with the metal. It follows that the rate of oxidation, for these strong fields, is determined only by the rate at which ions escape from the metal. This we must now calculate.

The potential energy of an ion in the surface layer of the metal is plotted in Figure 12. The diagram is intended to describe the state of affairs for an ion in the position P of Figure 2, that is, an ion ready to move into the oxide as the surface layer of metal dissolves. P represents the energy of the ion at rest at

this point; $Q_1, Q_2 \ldots$ are interstitial positions in the oxide and $S_1, S_2 \ldots$ the tops of the potential barriers separating these points. The heat of solution W_1 for a positive ion and the activation energy for diffusion U are shown in the diagram. It is convenient to set $W = W_1 + U$.

Then the chance per unit time that the atom will escape over the barrier to Q_1 is, in the absence of a field $\nu \exp(-W/\boldsymbol{k}T)$. In the presence of the field it is

$$\nu \exp(-W/\boldsymbol{k}T) \exp(qa'F/\boldsymbol{k}T),$$

where a' is the distance from P to the top of the barrier. The rate of growth of the film is thus

$$\frac{dX}{dt} = N'\Omega\nu \exp(-W/\boldsymbol{k}T) \exp\left(\frac{qa'F}{\boldsymbol{k}T}\right), \qquad \ldots\ldots(17)$$

where N' is the number of ions per unit area surface in sites such as P (Figure 2). This may be contrasted with the similar formula for weak fields (cf. equation (14)),

$$\frac{dX}{dt} = \frac{n_1 D_1 \Omega q F}{\boldsymbol{k}T}.$$

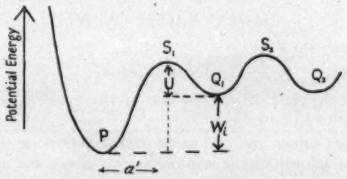


Figure 12. Potential energy of an interstitial ion in the neighbourhood of the metal–oxide interface.

Since $n_1 = a^{-3} \exp(-W_1/\boldsymbol{k}T)$, $D_1 = \nu a^2 \exp(-U/\boldsymbol{k}T)$, the two formulae could be comprised in the general formula

$$\frac{dX}{dt} = a\nu \exp(-W/\boldsymbol{k}T) \sinh\frac{qa'F}{\boldsymbol{k}T}, \qquad \ldots\ldots(18)$$

but for the fact that N' may be much less than $1/a^2$. A proper estimate of N' has not been made, since it depends on the number of "kinks" (Figure 2) present on the surface (Burton and Cabrera 1949); it is hoped to publish an estimate soon (Cabrera 1949 b).

Formula (17), for the growth rate in oxygen, can be used to describe two different phenomena:

(*a*) The growth of films in oxygen, where F is V/X, V being the contact potential difference between the metal and the adsorbed oxygen layer. This is the phenomenon discussed hitherto.

(*b*) The anodic formation of oxide films in an electrolyte containing oxide ions. V is then the voltage across the film.

In either case the equation may be written

$$dX/dt = u \exp(X_1/X), \qquad \ldots\ldots(19)$$

where $X_1 = qa'V/\boldsymbol{k}T$, $u = u_0 \exp(-W/\boldsymbol{k}T)$, $u_0 = N'\Omega\nu$.

X_1 is of order $10^{-6} - 10^{-5}$ cm., $u_0 \sim 10^4$ cm/sec. (or less). The formula (19) is valid only for $X \ll X_1$; it shows that the growth rate is very large for small X.

A particular and important consequence of formula (19) is that for constant V growth up to a certain limiting thickness occurs even at low temperatures where u is negligibly small; this will be the case when the solubility of metal ions in the oxide is negligible. This may be shown as follows. Suppose we say that growth has virtually stopped when one layer of atoms is added in 10^5 seconds, so that $dX/dt = 10^{-13}$. This occurs at a thickness X_L for which $\exp(X_L/X_1 - W/\boldsymbol{k}T) = 10^{-17}$. Substituting for X_1, this gives, since $17 \ln 10 = 39$,

$$X_L = Va'q/(W - 39\boldsymbol{k}T). \qquad \ldots\ldots(20)$$

Thus there exists a critical temperature $W/39\boldsymbol{k}$; for temperatures below this critical temperature the film grows rapidly up to some critical thickness and then stops, while for higher temperatures there is no limiting thickness, the initial rapid growth rate going over into the parabolic type of growth.

For $X \ll X_1$ an approximate integration of equation (19) can be given. Since

$$t = \frac{1}{u}\int_0^X \exp(-X_1/x)\,dx,$$

an integration by parts and neglect of higher terms in X/X_1 gives

$$ut = (X^2/X_1)\exp(-X_1/X).$$

For $X \ll X_1$ we may thus set

$$X_1/X = \ln(X_1 ut/X_L^2), \qquad \ldots\ldots(21)$$

giving a logarithmic type of growth law of the type $X_1/X = A - \ln t$. X_1 has values between 10^{-6} and 10^{-5} cm.

4.1. *Numerical values for an oxide which forms an excess semiconductor (aluminium).* For aluminium it is possible to deduce the unknown parameters W, a' in the theory from experimental work on the formation of anodic films on oxides. Following Verwey (1935) and Mott (1947a), we make use of experimental work by Gunterschultze and Betz (1934). These authors find that the current J through an oxide film during anodic formation depends on the field F through the formula

$$J = \alpha e^{\beta F}, \qquad \ldots\ldots(22)$$

where at room temperature $1/\alpha = 2{\cdot}75 \times 10^{16}$ μa/cm^2, or 0·92 E.S.U. If F is in volts, $\beta = 4{\cdot}2 \times 10^{-6}$. For small fields ($\beta F \sim 1$) the current is negligible. The formula is obviously to be compared with (17). We set

$$qa'/\boldsymbol{k}T = \beta, \qquad W = \boldsymbol{k}T \ln(N'qv/\alpha).$$

Taking $N' \sim 10^{15}$ cm^{-2}, $q = 3\boldsymbol{e}$, $\nu = 10^{12}$ sec^{-1} and $\boldsymbol{k}T = 0{\cdot}025$ e.v., we find $W = 1{\cdot}8$ e.v., $a = 3{\cdot}5 \times 10^{-8}$ cm. These values seem reasonable.

Inserting these values into formula (20) for the limiting thickness, we find for aluminium

$$X_L = 6 \times 10^{-8} V/(1 - T/530) \text{ cm.},$$

where V is measured in volts. At room temperature this gives, in cm., $10^{-7}V$.

At room temperature the thickness found by Cabrera and Hamon (1947) is about 20 A. (Figure 9), giving $V \sim 2$ volts. The same authors have been able to verify roughly the temperature dependence of the limiting thickness X_L given by the theory. They find that the limiting thickness increases slowly with temperature, while above 300° C. (573° K.) the growth is rapid and seems to continue without limit. This agrees well with the predicted value.

Cabrera, Terrien and Hamon (1947) have also found that an increase in the thickness of the order 50% could be produced by ultra-violet illumination (Figure 9). This has been explained by Cabrera (1949a) as due to the ejection of electrons from the metal to the adsorbed oxygen, thereby increasing the field in the layer. This is only possible when the thickness of oxide is bigger than 10 A., otherwise the current of electrons going from the absorbed oxygen back to the metal by tunnel effect will compensate the increase in the electrons going in the reverse direction produced by photoelectric effect. This is the behaviour actually observed (Figure 9).

4.2. *Oxides which absorb oxygen*; *the case of copper*. In the case of Cu_2O forming on copper the difference of potential set up across the oxide film must be 0·7 volts. The energy levels of Cu_2O are as shown in Figure 13. In the first place the interval from the full band of Cu_2O to the top of the Fermi distribution in the metal is known to be 0·7 ev., this being the height of the Schottky barrier in a $Cu-Cu_2O$ rectifier (Mott and Gurney 1948, p. 189). In the second place the full band is believed to be due to the Cu^+ 3d shells, and will probably be *higher* than the empty levels due to the adsorbed oxygen. Thus in equilibrium charge will distribute itself as in Figure 13 (b).

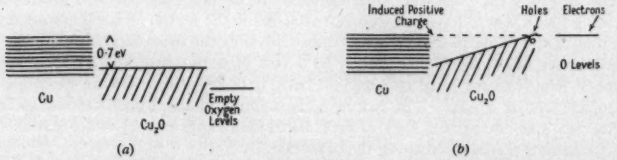


Figure 13. Electronic levels in Cu, Cu_2O and oxygen: (*a*) before equilibrium is reached, (*b*) afterwards.

The critical process which is so much accelerated by the field must now be the formation of a vacant cation site at the oxide–air interface. Though the mechanism must be very similar to that described for aluminium, the details have not yet been worked out.

The fact that the thickness of oxide formed at room temperature (130 A.) is much bigger than that formed on aluminium, in spite of the fact that V is smaller, could be explained by the fact that the critical temperature $W/39k$ at which the logarithmic law goes over into a cubic law is now just above room temperature, if we take for $W=W_1+U$ the value 1·1 ev. suggested by the measurements of Campbell and Thomas. At liquid air temperature we should expect a thinner layer on Cu than Al, but this does not seem to be the case according to new measurements now in progress at this Laboratory by Aitchison and Allen.

§5. ADHESION AND CRYSTAL FORM OF AN OXIDE FILM

We do not know of any detailed theoretical discussion of the cohesive forces between an oxide layer and a metal substrate, or any experimental measurement of the surface energy of the interface. However, all oxides are at least partly polar; and the charges on the metal and oxygen ions must be strongly attracted to the substrate metal. Thus strong cohesive forces between metal and oxide must exist, whether or not the lattice parameters of the metal and of the oxide are equal or nearly equal.

Nevertheless, the influence of the lattice parameter of the substrate metal on the structure of the oxide is of great importance. In some cases (ZnO on the base planes of Zn) a pseudomorphic structure grows up to a considerable thickness (Finch and Quarrell 1934, 1939), the normal oxide structure being squashed into congruence with that of the zinc underneath. In other cases an oriented overgrowth is observed (Cu_2O, FeO); in others (e.g. Al_2O_3 on Al) no orientation is apparent.

A theoretical discussion of oriented overgrowths has recently been given by Frank and van der Merwe (1949, see also van der Merwe 1949) which promises to be of great importance for the theory of oxidation. These authors consider what will happen when the first monolayer of a film to be deposited on a metal is put down. Their theory is applicable first of all to one solid put down by evaporation or electro-plating on another, but their ideas are useful also for a metal exposed to oxygen. In this case, as soon as a monolayer of adsorbed oxygen is formed, atoms from the metal will start to pass through it by the mechanism which has already been described. By a monolayer of oxide, then, we mean a monolayer of oxygen through which one layer of metal atoms has passed.

Frank and van der Merwe denote by a the lattice spacing of the substrate metal and by b the "natural" spacing of the oxide layer, by which is meant the spacing that it would have if it were not attached to the metal. This is assumed to be that of the oxide in bulk. They then represent the monolayer of oxide (in a preliminary one-dimensional model) by a row of balls connected by springs of natural length b and force constant μ; that is to say, the restoring force for a displacement δx is $\mu\delta x$. The row of balls is acted on also by a force due to the substrate, the potential energy of each ball being represented, as a function of its coordinate x measured along the substrate, by

$$\tfrac{1}{2}W\cos(2\pi x/a). \qquad \ldots\ldots(23)$$

The "misfit" M between the substrate and the monolayer is defined by $M=(b/a)-1$. In practical cases b is greater than a and M is positive. The fit or misfit of the monolayer and substrate is naturally described in terms of dislocations. If, for instance, 99 or 101 atoms (balls) lie over 100 of the troughs in the potential (23), then in equilibrium the majority of the atoms lie nearly at the bottoms of their troughs, while there is a small region where the atoms ride over the crests, to miss a trough or squeeze an extra atom in. This region of misfit is called a surface dislocation; if a crystal is built above it, it will develop into a dislocation of the type used in the theory of plastic flow (Taylor 1934; for a review, see Bristol Conferences on Solids, 1940 and 1947, published by the Physical Society).

The mathematical investigation shows that if a line of balls (atoms) of finite length is put down on the substrate, then *in its state of lowest energy* each ball will lie at the bottom of a trough, except just near the ends, so long as the misfit M is less than a critical misfit M_0 given by

$$M_0=\frac{2}{\pi}\left(\frac{\mu a^2}{2W}\right)^{-\frac{1}{2}}.$$

If M exceeds M_0 by more than a very small proportion, however, and the chain is in its state of lowest energy, the spacing that the chain will take up is very nearly its equilibrium value b. However, even if the misfit is greater than M_0, then, if a line of balls is put down with each ball at the bottom of a trough, they

will remain there in metastable equilibrium unless M is greater than $\frac{1}{2}\pi M_0$—some 40% greater. Only if M is greater than $\frac{1}{2}\pi M_0$ will dislocations be generated spontaneously at the edge of the chain, so that a chain put down in a compressed state (as will happen if the balls are put down one by one) can expand to its natural length.

The application of this model to oxidation is as follows: an estimate of the interatomic forces shows that M_0 may be of the order 0·09 in an average case for a monolayer of oxide, though there may be wide variations from one metal to another. Thus, if the misfit is less than 9%, the first monolayer of oxide should have the lattice parameter of the substrate. If the temperature of deposition is low, the same will be true up to a misfit of 14%; but at high temperatures dislocations may be generated at the edge of the layer, so the value of the misfit, below which the oxide layer will in fact take up the parameter of the substrate, is smaller. If the layer does *not* take up the parameter of the substrate, each island of oxide would rotate very easily about a line perpendicular to the surface; we should not in this case expect any orientation, either in the first monolayer or in subsequent growth, except perhaps for a common crystal axis.

Suppose then that the degree of misfit is less than 9 or 14%, or whatever the figure may be, and that the first monolayer of oxide grows with the lattice parameter of the substrate and covers the whole available surface. By that time the second and third layers will have begun their growth, and will soon cover the whole surface too. The film, compressed as it is to fit the lattice parameter of the substrate, is no longer in its state of lowest energy; in terms of the ball and spring model, μ has become two or three times as big without any great change in W, it being assumed that W is mainly due to interaction between nearest neighbours. But none the less the film is in metastable equilibrium; it cannot expand by forming dislocations at the edges, because if the film covers the surface there are no free edges.

We believe that films such as that of ZnO on zinc are formed in this way. The first monolayer of oxide is formed with the parameter of the substrate, and the film continues to grow with this parameter. In the case of aluminium, on the other hand, the initial misfit is too great for this to happen, and a polycrystalline or amorphous film results.

It seems *a priori* highly probable that the thin films (*c.* 100 A.) formed at moderately low temperatures by the mechanism described in §4 are compressed to fit the substrate in the way described here, and remain in that state. The compressive strains, of the order 10%, are, of course, much greater than the bulk material can support; in zinc the misfit is actually 20%; but it is a fairly obvious consequence of the modern theory of strength of solids (see, for instance, Mott 1949 b) that very thin films should show a much higher compressive strength than the bulk material. At some period in the growth, however, the film must break away and achieve its equilibrium lattice parameter; this will certainly have occurred for thick films (*c.* 10^{-4} cm.) growing according to the mechanism of §2. Two mechanisms are possible by which the film may break away: slip or recrystallization. One or other of those processes is probably responsible for the kink in the oxidation curve shown for copper in Mrs. Garforth's work illustrated in Figure 1.

Some of the consequences of any mechanism by which the film continually breaks and heals have been explored by Evans, who in particular has shown that a logarithmic growth law is to be expected in cases of this sort; and when growth to a thickness of several hundred Ångströms is observed to follow a logarithmic law, it seems very probable that a mechanism such as this is valid. In general we may say that this intermediate region between the very thin, probably pseudomorphic films of 100 A. or less in thickness, described in §4, and the thick films growing according to the parabolic law is imperfectly understood.

In conclusion, we would like to express our thanks to Dr. J. W. Mitchell and the group working under his direction on the experimental side of this subject for many discussions and permission to reproduce some of their results prior to publication.

REFERENCES

BARDEEN, J., 1947, *Phys. Rev.*, **71**, 374.
BARDEEN, J., BRATTAIN, W. H., and SHOCKLEY, W., 1946, *J. Chem. Phys.*, **14**, 714.
BATES, D. R., and MASSEY, H. S. W., 1943, *Phil. Trans.*, **239**, 269.
BÉNARD, J., and COQUELLE, O., 1946, *C.R. Acad. Sci., Paris.*, **222**, 796.
BÉNARD, J., and TALBOT, J., 1948, *C.R. Acad. Sci., Paris*, **225**, 411.
BURTON, W. K., CABRERA, N., and FRANK, F. C., 1949, *Nature, Lond.*, **163**, 398.
BURTON, W. K., and CABRERA, N., 1949, *Trans. Faraday Soc.*, in press.
CABRERA, N., 1949 a, *Phil. Mag.*, **40**, 175; 1949 b, to be published.
CABRERA, N., and HAMON, J., 1947, *C.R. Acad. Sci., Paris*, **224**, 1713.
CABRERA, N., TERRIEN, J., and HAMON, J., 1947, *C.R. Acad. Sci., Paris*, **224**, 1558.
CAMPBELL, W. E., and THOMAS, U. B., 1947, *Trans. Electrochem. Soc.*, **91**, 345.
EVANS, U. R., 1946, *Nature, Lond.*, **157**, 732; 1948, *Metallic Corrosion*, 2nd ed. (London. Edward Arnold).
FINCH, G. I., and QUARRELL, A. G., 1933, *Proc. Roy. Soc.* A, **141**, 398; 1934, *Proc. Phys. Soc.*, **46**, 148.
FRANK, F. C., 1949, *Trans. Faraday Soc.*, in press.
FRANK, F. C., and VAN DER MERWE, J., 1949 a, *Nature, Lond.*, in press; 1949 b, *Proc. Roy. Soc.* A, in press.
GARFORTH, F., 1949, to be published.
GULBRANSEN, E. A., and WYSONG, W. S., 1947, *J. Phys. Colloid. Chem.*, **51**, 1087.
GUNTERSCHULTZE, A., and BETZ, H., 1934, *Z. Phys.*, **92**, 367.
GWATHMEY, A. T., and BENTON, A. F., 1940, *J. Chem. Phys.*, **8**, 431; 1942, *J. Phys. Chem.*, **46**, 969.
KRUPKOWSKI, A., and BALICKI, S., 1937, *Métaux et Corrosion*, **12**, 89.
LENNARD-JONES, J. E., and DENT, B., 1928, *Trans. Faraday Soc.*, **24**, 92.
VAN DER MERWE, J., 1949, *Trans. Faraday Soc.*, in press.
MOTT, N. F., 1940, *Trans. Faraday Soc.*, **39**, 472; 1947 a, *Ibid.*, **43**, 429; 1947 b, *J. Chim. Phys.*, **44**, 172; 1949 a, *J. Instn. Elect. Engrs.*, in press; 1949 b, *Research*, **2**, 162.
MOTT, N. F., and GURNEY, R. W., 1948, *Electronic Processes in Ionic Crystals*, 2nd ed. (Oxford: University Press).
STEINHEIL, A., 1934, *Ann. Phys., Lpz.*, **19**, 465.
TAYLOR, G. I., 1934, *Proc. Roy. Soc.* A, **145**, 362.
VERWEY, E. J. W., 1935, *Physica*, **2**, 1059.
WAGNER, C., and GRÜNEWALD, K., 1938, *Z. Phys. Chem.* (B), **40**, 455.

FRACTURE AND STRENGTH OF SOLIDS

By E. OROWAN
Cavendish Laboratory, Cambridge

CONTENTS

I. Introduction — Page
§ 1. General survey 186
§ 2. The main types of fracture 186

II. Ultimate stress and necking
§ 3. Ultimate stress 188
§ 4. The necking of tensile specimens 189
§ 5. Stability of the extension of Newtonian viscous materials 190
§ 6. Stability of the extension of visco-plastic materials 191

III. Brittle fracture
§ 7. Critical tensile stress law 191
§ 8. Molecular cohesion (" theoretical strength ") 192
§ 9. The Griffith theory of brittle strength 192
§ 10. Derivation of the Griffith fracture condition from atomic considerations 194
§ 11. Experimental verification of the Griffith theory 194
§ 12. Length of the Griffith cracks; means of observation; mechanisms of crack formation 196
§ 13. Statistics of cracks 197
§ 14. Size effect 198
§ 15. Adsorption effects 198
§ 16. The static fatigue of glass 199
§ 17. Brittle fracture under bi- and triaxial stress 199
§ 18. Brittle strength of polycrystalline aggregates 202

IV. Ductile fractures and notch brittleness
§ 19. Rupture by necking and by slipping-off 202
§ 20. General features of ductile fracture 203
§ 21. The tensile fracture of ductile metals 204
§ 22. The " true " tensile strength 205
§ 23. The Ludwik hypothesis 206
§ 24. Notch brittleness 207
§ 25. Criterion for fully ductile, notch brittle and brittle behaviour 209
§ 26. The transition temperature between full ductility and notch brittleness 211
§ 27. Notch brittleness and low temperature brittleness 212
§ 28. Temper brittleness 213
§ 29. Size effect in notch brittleness 213
§ 30. Kuntze's " technical cohesive strength " 214
§ 31. Non-existence of a tensile stress criterion for fibrous fracture 216
§ 32. Other suggested criteria for ductile fracture 218
§ 33. Ductile fracture as a crack propagation process 218
§ 34. Two special mechanisms of ductile fracture 219

V. Fatigue
§ 35. Mechanical fatigue under cyclic stressing 221
§ 36. The nature of fatigue fracture 222
§ 37. Theory of mechanical fatigue 223
§ 38. Corrosion fatigue and stress corrosion 228

VI. Intercrystalline fractures
§ 39. Intercrystalline viscous fracture (creep fracture) 228
§ 40. Intercrystalline brittle fracture 229

VII. Fracture by molecular sliding and special types of fracture
§ 41. Fracture by molecular sliding 230
§ 42. Special types of fracture 230

References.

I. INTRODUCTION

§1. GENERAL SURVEY

STRENGTH is the resistance of a material to fracture; in the quantitative sense, it is a critical value of stress at which fracture occurs. In a loose usage which has not completely disappeared yet from the engineering and metallurgical literature, strength denotes the stress at which the material "fails" either by fracture or by plastic deformation. However, the stress at which plastic deformation takes place should be called the yield stress.

Fracture is not one single physical phenomenon; there are several essentially different processes that may lead to the disintegration of a body by the action of mechanical forces. Much of the confusion that existed and still exists in the physical theory of strength is due to insufficient attention to this circumstance. A review of fracture phenomena, therefore, must begin with an enumeration of the main types of fracture; in the following sections, the corresponding fracture mechanisms will be discussed.

The same material may fail with different mechanisms of fracture according to the stress and strain conditions and to the temperature; thus, a low carbon steel may show the fibrous and shear types of fracture at room temperature, brittle fracture below $-80°$ C., and intergranular creep fracture in slow straining at temperatures above 600° C. Consequently, a material may have several characteristic strength values if several of its possible fracture mechanisms operate at definite critical values of certain stress components. In anisotropic materials, the strength for a given type of fracture may depend on the direction of the stress; thus, crystals have in general different strengths for brittle fracture along crystallographically different sets of cleavage planes, and the operative cleavage plane is then determined by the orientation of the crystal relative to the principal stress axes.

Under triaxial stress, the fracture condition will in general involve all principal stresses; the behaviour of the material cannot be described then by a single fracture stress. Nevertheless, the equation expressing the fracture condition may contain only one physical constant, the value of which then determines the resistance of the material to fracture. This seems to be the case in brittle fracture (cf. §17).

With plastic materials, the fracture condition involves the entire strain history in addition to the principal stresses and (if the material is anisotropic) the orientation of the principal axes. Many authors assume that, at least approximately, the fracture behaviour of ductile materials can be described by two curves: one giving the brittle strength, the other the ductile strength as functions of the amounts of deformation. It will be seen, however (cf. §31), that ductile fractures do not obey a simple critical stress condition; one cannot, therefore, speak of a definite ductile strength of a material. This is most clearly seen in the case of intergranular creep fracture which is due to grain boundary viscosity. Any stress within a wide range can cause fracture in a high temperature creep test: the time needed for fracture is short or long according to whether the stress is high or low.

§2. THE MAIN TYPES OF FRACTURE

(i) *Brittle fracture* (cleavage fracture).—This is the theoretically best known mechanism of fracture. It is the only one that occurs in completely brittle

materials; however, it is also observed in ductile materials under certain conditions (cf. the notch brittleness of steel).

(ii) *Rupture* (by localization of plastic deformation).—Many very ductile metals, particularly in sheet form, suffer rupture in tension by continued thinning around one section ("necking"): the parts above and below the neck finally separate along a sharp edge or in a point, like a thread of treacle or hot glass. Similarly, single crystals of ductile metals may separate into two parts, when strained slowly at a sufficiently high temperature, by slip confined to one or a few slip planes which ultimately results in two parts of the crystal slipping off one another.

Such processes cannot well be called fracture; in what follows, they will be called rupture.

(iii) *Fibrous fracture.*—This forms the bottom of the cup in the cup-and-cone fracture of ductile metals in the tensile test. The surface of fracture appears velvety matt to the naked eye, and deeply jagged under the microscope; it is approximately perpendicular to the greatest tension.

(iv) *Shear fracture.*—This produces the side of the cup in the cup-and-cone fracture; it is the typical fracture of ductile metals in shear and torsion, and in tensile tests on sheets. The fracture follows a surface of maximum shear strain; the surface of fracture is often remarkably smooth.

(v) *Fatigue fracture.*—Fracture by mechanical fatigue under cyclic stressing has many similarities to brittle fracture, for example, in its sensitivity to surface defects, and in that the surface of fracture is usually perpendicular to the greatest tension. However, typical fatigue fracture cannot occur without local plastic deformation at the tip of the propagating crack, and a thin layer at the surface of fracture is intensely cold worked.

Corrosion fatigue, stress corrosion, and static fatigue phenomena in brittle materials (e.g. the delayed fracture of glass) have different mechanisms.

(vi) *Intergranular viscous fracture* (creep fracture).—This occurs only if the temperature is high and the rate of deformation low enough for the grains to slide over one another without considerable intracrystalline deformation. As a consequence of the sliding, cavities open up between the grains and finally fracture occurs.

(vii) *Intergranular brittle fracture.*—In numerous cases, impurities segregating or accumulating along the grain boundaries in metals lower the cohesion between the grains so much that fracture can occur by intergranular separation with little or no plastic deformation even if the individual grains are very ductile.

(viii) *Fracture by molecular sliding.*—This is similar to the intergranular viscous fracture in that the structure of the material becomes loosened by a viscous type of sliding between structural elements; in the present case, however, the elements are not grains of polycrystalline metals but large molecules of organic materials. Fracture is often preceded by the development of opacity due to the opening up of intermolecular cavities.

(ix) *Special types of fracture.*—Each of the types enumerated above occurs with many different materials; in addition, there are numerous mechanisms of fracture that arise from some special structural or textural feature confined to a rather special class of materials. Thus, for example, materials with a dendritic texture often break along the dendrite boundaries (Iitaka 1931, Iitaka and

Yamagishi 1938), and twinned or kinked crystals along the planes of twinning or kinking (Orowan 1942). Some of the special mechanisms are closely related to one of the more general types; others, however, are too individual to be forced into a general scheme.

II. ULTIMATE STRESS AND NECKING

§3. ULTIMATE STRESS

Before the various mechanisms of fracture and the factors determining strength are considered, the "ultimate stress" or "ultimate tensile strength" of ductile materials should be discussed. This is a quantity of great practical importance, which, however, has nothing to do with fracture. It is defined as the maximum load reached in a tensile test, divided by the initial cross-sectional area of the specimen; that is, it is the maximum "conventional" stress (the load divided by the initial cross-sectional area of the specimen is often called the conventional stress). The ultimate stress can be determined from the true stress–strain curve

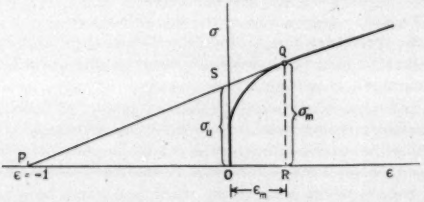


Figure 1. Geometrical construction of the ultimate stress from the true stress–strain curve.

of the material by a simple geometrical construction, without any reference to fracture and strength (Considère 1885). Let σ be the tensile stress and ϵ the linear strain defined by

$$\epsilon = (x - x_0)/x_0, \qquad \ldots\ldots(1)$$

where x_0 is the initial and x the current cross-sectional area. The tensile force $F = a\sigma$ (a = cross-sectional area) reaches a maximum when

$$dF = \sigma\, da + a\, d\sigma = 0, \quad \text{or} \quad da/a = -d\sigma/\sigma. \qquad \ldots\ldots(2)$$

Since the volume $V = ax$ remains practically constant during plastic extension,

$$da/a = -dx/x; \qquad \ldots\ldots(3)$$

combining this with (2), one obtains

$$d\sigma/\sigma = dx/x \qquad \ldots\ldots(4)$$

as the condition for the load maximum. In view of the definition (1) of the strain, $dx/x = (dx/x_0)(x_0/x) = d\epsilon/(1+\epsilon)$; hence, (4) can be written as

$$\frac{d\sigma}{d\epsilon} = \frac{\sigma}{1+\epsilon}. \qquad \ldots\ldots(5)$$

This condition of the maximum load can be expressed by a simple graphical construction in the true stress–strain diagram (Figure 1). Let P be the point on the abscissa axis whose abscissa is -1. For any point Q on the stress–strain

curve $\sigma(\epsilon)$, the right hand side of (5) is the gradient of the straight line connecting P with Q. The condition of the load maximum is satisfied if this gradient is equal to the gradient $d\sigma/d\epsilon$ of the curve at the point Q; that is, the load maximum occurs at the point of contact of the curve with its tangent drawn from the point P. $RQ = \sigma_m$ is the true stress at the maximum load point, and $OR = \epsilon_m$ the maximum load strain. If F_m is the maximum load and so $\sigma_m = F_m/a_m$, the ultimate stress is, in view of (1) and (3), $\sigma_u = F_m/a_0 = (F_m/a_m)(a_m/a_0) = \sigma_m(x_0/x_m) = \sigma_m/(1+\epsilon_m)$, i.e.

$$\sigma_u/\sigma_m = 1/(1+\epsilon_m). \qquad \ldots\ldots(6)$$

The ultimate stress σ_u, therefore, is the intercept OS on the stress axis between the strain axis and the tangent PQ.

Corresponding graphical constructions can be obtained if the yield stress is given as a function, not of the linear strain, but of the logarithmic strain $\ln(x/x_0)$, or of the reduction of area $(a_0-a)/a_0 = (x-x_0)/x$.

According to equation (5) and Figure 1, the ultimate stress is entirely determined by the stress–strain curve, that is, by the plastic behaviour of the material, without any reference to its strength properties, provided that fracture does not occur before the load maximum corresponding to equation (5) is reached. If premature fracture does occur, the highest load observed is a genuine expression of the resistance to fracture. The values of the "tensile strength" found in data books may be either genuine strengths referred to the initial cross section, or conventional maximum load stresses unconnected with strength, according as the material is rather brittle or rather ductile; if the maximum reduction of area is given, a decision between these alternatives is often, but not always, possible. In order to avoid ambiguities, plastic maximum load stresses should never be called strength, and ultimate stresses and fracture stresses should be tabulated separately. For reasons to be discussed further below, true tensile fracture stresses for very ductile materials like Al, Au, Cu, or Ni have never been measured; as already indicated, ductile fracture of the fibrous type does not obey a critical stress law and so for such fractures a strength in the strict sense of the word does not exist.

The occurrence of a plastic load maximum, of course, is peculiar to tensile loading. There is no quantity corresponding to the ultimate stress in compression or shear. If a load maximum is observed in such tests (a rare occurrence), it must be due to a true softening or loosening of the material with increasing deformation. The most common example of such behaviour is the yield phenomenon of many ferritic steels; with organic materials, numerous examples of strain softening ("thixotropy") are known.

§4. THE NECKING OF TENSILE SPECIMENS

Before the maximum load is reached, the deformation of a tensile specimen is stable; if one cross section is slightly weaker than the rest, it suffers a slight additional deformation which raises its yield load by strain hardening to the yield load of the normal cross sections. When the maximum load has been exceeded, however, a slight additional deformation causes a decrease, instead of an increase, of the yield load; a weaker cross section becomes thereby weaker still, the deformation remains confined to its neighbourhood, and ceases everywhere else in the specimen. Since there is always a cross section that is slightly

thinner or weaker than the others, necking occurs generally after the maximum load has been reached. After necking has begun, the strain in the uniformly extended parts outside the region of the neck remains constant at the value $\epsilon_m = OR$ (cf. Figure 1) of the maximum load strain. For this reason, ϵ_m is usually called the " uniform strain ".

§ 5. STABILITY OF THE EXTENSION OF NEWTONIAN VISCOUS MATERIALS

The considerations of the preceding two sections refer to purely plastic materials whose law of deformation is given by a stress–strain curve. Such a curve represents sufficiently accurately the plastic behaviour of metals with melting points above, say, 600° if they are extended not too slowly at room temperature. In creep tests at elevated temperatures, however, metals behave more like viscous materials whose law of deformation is a functional relationship between the stress and the rate of deformation. The question arises then whether the instability leading to the formation of the neck occurs with viscous materials also.

This can be answered very simply (Nadai and Manjoine 1941, Orowan 1949a) for Newtonian viscous materials whose law of deformation is

$$\sigma = 3\eta \frac{d\epsilon}{dt}, \qquad \ldots\ldots(7)$$

where η is the coefficient of viscosity and t the time; $d\epsilon$ is the differential of the logarithmic strain, or, what is the same, of the linear strain referred to the length of the specimen at the moment considered: $d\epsilon = dx/x$. Since the volume remains practically constant, equation (3) gives $d\epsilon = -da/a$; further, $\sigma = F/a$ if, again, F is the tensile force and a the current cross-sectional area. If these values are introduced in (7), one obtains after integration

$$a = a_0 - Ft/3\eta. \qquad \ldots\ldots(8)$$

During extension by a constant force, therefore, the cross section of a Newtonian viscous specimen decreases by constant amounts per unit of time; the specimen would extend to infinite length within the finite time $3\eta a_0/F$ if its material could withstand infinite stresses without fracture.

Let it be assumed now that the specimen had a weak point where the initial cross-sectional area a_0' was slightly less than a_0. After a time t, this cross-sectional area would have decreased to

$$a' = a_0' - Ft/3\eta. \qquad \ldots\ldots(8a)$$

Subtraction of (8a) from (8) gives

$$a - a' = a_0 - a_0'. \qquad \ldots\ldots(9)$$

The difference between the normal and the deficient sectional areas, therefore, remains constant during the extension; there is no tendency to necking until the cross-sectional areas have decreased to the order of magnitude of the initial difference between the normal and the deficient sectional areas. In other words, the extension of a Newtonian viscous specimen does not show the instability manifested by the purely plastic specimens. This is the main reason why glass rods and tubes can be made by pulling from hot glass, while metals cannot be formed in this way.

§6. STABILITY OF THE EXTENSION OF VISCO-PLASTIC MATERIALS

Metals are neither purely plastic nor purely viscous; it is of interest, therefore, to investigate the behaviour of a material of an intermediate nature. The simplest type of visco-plastic material has the law of deformation shown in Figure 2:

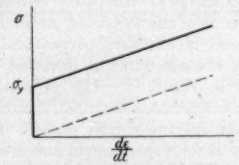


Figure 2. Law of deformation of a viscous material with yield point ("Bingham solid"). Dashed line: Newtonian viscosity.

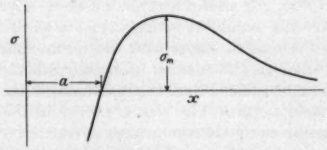


Figure 3. Cohesive forces as a function of the interatomic spacing.

there is a yield point, and, after exceeding this, the stress is a linear function of the strain rate ("Bingham material"). In the same graph, the dashed line would represent Newtonian viscous behaviour.

The stability investigation for a Bingham material is quite similar to that of the preceding section (Orowan 1949 a); the qualitative result, however, is clear without calculation. At high rates of deformation, the stress is nearly proportional to the deformation rate, and the existence of a yield point becomes irrelevant. The specimen behaves then like a Newtonian one, and its extension is stable. At very low rates of extension, on the other hand, the stress is only slightly higher than the yield stress, and its dependence upon the extension rate can be neglected; the material behaves then like a plastic one with a sharp yield point, and necking will occur at the beginning of the extension.

III. BRITTLE FRACTURE

§7. CRITICAL TENSILE STRESS LAW

Experience shows that isotropic brittle materials break in uniaxial tension when the tensile stress reaches a critical value, that of the *brittle* (tensile) *strength*. The strength of very thin fibres usually increases with decreasing thickness; this effect will be discussed in §14.

The tensile fracture surface of crystals is frequently a crystallographic plane of low index and high atomic density ("cleavage plane"); in many crystals, however (e.g. quartz), the surface of fracture is non-crystallographic and of the same appearance as in glass. If the surface of fracture is crystallographic, the fracture obeys, in general, Sohncke's law (Sohncke 1869): it occurs when the tensile stress normal to the operative cleavage plane reaches a critical value. In alkali halides, the cleavage plane is (100), and the critical normal stress of the order of 20 bars *.

The fracture condition for triaxial states of stress will be discussed in §17.

* 1 bar = 10^6 dynes/cm^2.

§8. MOLECULAR COHESION ("THEORETICAL STRENGTH")

It was recognized at an early stage that the tensile strength of a perfectly homogeneous and flawless brittle solid ought to be of the order of its intrinsic pressure for which, in the case of water, a good estimate has already been given by Thomas Young (1805). The simplest way of obtaining a rough estimate of the molecular cohesion of a solid is from its surface energy (Polanyi 1921, Orowan 1934 a). If a solid is extended uniformly, the excess of the molecular attraction over the repulsion acting across a surface element of unit area perpendicular to the tension varies with the mean interatomic distance x in the direction of tension in the manner indicated schematically in Figure 3. The extension should be stable until the maximum of the curve is reached; after that, fracture should occur. The elastic energy in the stressed specimen must provide the surface energy of the surfaces of fracture which is 2α for a specimen of unit cross section (α is the specific surface energy). It is clear that a large fraction (perhaps not much less than a half) of the energy 2α must be present at the moment of fracture between the molecules immediately adjacent to the surface of fracture; in other words, if σ_m is the stress at which the fracture begins and thus $\sigma_m^2/2E$ the density of elastic energy if Hooke's law with the usual value of Young's modulus E is assumed to hold up to fracture, the energy $a\sigma_m^2/2E$ present between two neighbouring atomic planes of spacing a must be of the order of magnitude of α.

Hence, the order of magnitude of the fracture stress σ_m must be

$$m = (2\alpha E/a)^{\frac{1}{2}}. \qquad \ldots\ldots (10)$$

For technical metals, stones, and glasses, E is of the order of 10^5 or 10^6 bars; the order of magnitude of α is 10^3 ergs/cm^2, and that of a, 3×10^{-8} cm. For such materials, therefore, σ_m must be of the order of 10^{11} dynes/cm$^2 = 10^5$ bars. Tensile stresses of this order of magnitude have been measured in freshly prepared glass rods (Griffith 1920), mica sheet with stressless edges (Orowan 1933 a) and thin tungsten wires; steel piano wires can withstand tensile stresses up to about 25,000 bars. However, such high strengths are rare; the tensile strength of ordinary window glass is below 1,000 bars, and that of alkali halides, as mentioned, not much higher than 20 bars. The cause of this discrepancy between the molecular cohesive forces and the observed tensile strengths has been elucidated by Griffith (1920, 1924); the following sections will deal with his work and its subsequent development.

The molecular cohesion of NaCl crystals for tensile straining in the (100) direction has been calculated by Zwicky (1923) from the Madelung–Born theory of ionic lattices; the result was a theoretical strength of about 20,000 bars. Almost the same figure would be obtained from the rough estimate given by equation (10).

§9. THE GRIFFITH THEORY OF BRITTLE STRENGTH

Griffith assumed that the discrepancy between the theoretically estimated and the observed values of the tensile strength was due to the presence of very small cracks or other flaws around which a strong stress concentration arose when the solid was stressed. According to his theory, the theoretical strength is the true microscopic fracture stress which is actually reached in very small volumes of the specimen while the mean stress may remain very low. In calculating the stress concentration at the tip of a crack of given length, Griffith made use of Inglis' calculation (Inglis 1913) of the stress distribution around an elliptical

hole in a stressed plate, by regarding the crack as a very flat elliptical hole. If a plate containing a flat elliptical hole of major axis $2c$ is subjected to a tensile stress σ perpendicular to the major axis, the highest tensile stress which occurs at the ends of the major axis is

$$\sigma_m = 2\sigma(c/\rho)^{\frac{1}{2}}, \qquad \ldots\ldots (11)$$

where ρ is the radius of curvature at the ends of the major axis. σ_m tends to infinity as ρ decreases to zero; without further physical assumptions, therefore, no definite value for σ_m can be obtained. Griffith circumvented the difficulty by assuming that the crack will lengthen and thus lead to fracture if, for a small increase of its length $2c$, the work of the external forces is sufficient to cover the increase of the elastic energy around the crack and of its surface energy. Surface cracks of depth c produce approximately the same stress concentration as internal cracks of length $2c$; in the present section, usually internal cracks will be considered, but the results are equally valid for surface cracks of length c. From the stress distribution around an elliptical hole, the excess elastic energy of the plate with a crack (above the energy of the same plate without a crack) can be calculated as

$$W_e = \pi c^2 \sigma^2 / E \qquad \ldots\ldots (12)$$

per unit of thickness if this is small compared with the length $2c$ of the crack, that is, if there is a plane state of stress. If the thickness is large compared with $2c$ (plane strain), the excess elastic energy is

$$W_e = (1-\nu^2)\pi c^2 \sigma^2 / E \qquad \ldots\ldots (12\,a)$$

where ν is Poisson's ratio. The work of the external forces when the crack is introduced is twice the excess elastic energy, W_e, and the surface energy of the crack is

$$W_s = 4\alpha c \qquad \ldots\ldots (13)$$

per unit of thickness of the plate. The crack is in (unstable) equilibrium with the external forces if, for a small increase of its length, the increase of the surface energy and of the excess elastic energy is just covered by the work done by the external forces; that is, if

$$\frac{d}{dc}(W_e - W_s) = \frac{2\pi c}{E}\sigma^2 - 4\alpha = 0 \quad \text{or} \quad \sigma = (2\alpha E/\pi c)^{\frac{1}{2}} \qquad \ldots\ldots (14)$$

for a thin plate; the corresponding condition for the thick plate is

$$\sigma = \{2\alpha E/\pi c(1-\nu^2)\}^{\frac{1}{2}}. \qquad \ldots\ldots (14\,a)$$

Since the equilibrium is unstable, the crack begins to run and fracture occurs as soon as the stress exceeds the value given by (14) or (14 *a*).

Sack (1946) and Elliott (1947) have extended the theory to the axially symmetrical three-dimensional case where the crack is a very flat oblate ellipsoid (" penny-shaped " crack). Sack's result is that in this case the expression for σ differs from (14 *a*) only in that the numerical factor under the square root is $\pi/2$ instead of $2/\pi$. Elliott has taken into consideration details of the interatomic forces acting across the plane of the crack; this will be discussed in the following section.

§10. DERIVATION OF THE GRIFFITH FRACTURE CONDITION FROM ATOMIC CONSIDERATIONS

Obviously, there is no physical sense in considering radii of curvature at the tip of a crack whose order of magnitude is less than the intermolecular distance a; the sharpest possible crack can only have a radius of curvature of the atomic order of magnitude a. If, in equation (11), ρ is replaced by a and the resulting equation divided by equation (10), one obtains for the value of the applied mean stress at which the microscopic stress at a crack of atomic sharpness reaches the value of the molecular cohesion (Orowan 1934a),

$$\sigma = (\alpha E/2c)^{\frac{1}{2}}. \qquad \ldots\ldots (15)$$

This differs from Griffith's equation (14) only by the factor under the square root being $\frac{1}{2}$ instead of $2/\pi$, a difference quite within the approximations used in the Griffith theory (e.g. the use of Hooke's law up to the moment of fracture).

Elliott (1947) has investigated the question of the errors introduced by the assumption of an elliptical crack. In a crystal with a cleavage plane, a crack must be of the shape shown in Figure 4 where the lines A–A and B–B indicate the

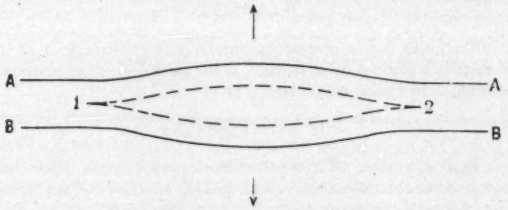


Figure 4. Cleavage crack in a crystal.

lattice planes nearest to the symmetry plane of the crack after the application of a tension perpendicular to the plane of the crack. In a solid continuum with a perfectly sharp crack whose profile, after applying the tension, is given by the dashed line in Figure 4, the stress would be infinitely high at the points 1 and 2; if the same stress distribution applies to the real solid, however, the stress remains finite everywhere in the nearest planes A–A and B–B that actually contain matter. Elliott showed that the relationship between the normal stress in the planes A–A and B–B and their spacing at different points, as calculated from the continuum model, happens to be very similar to what ought to be expected for the relationship between the interatomic forces acting across the crack at different points and the distance between the atomic planes A–A and B–B. A detailed calculation has shown that the cleavage type of crack gives the same fracture condition, with only slightly different values of the numerical factors, as the elliptical crack of Griffith and the ellipsoidal (penny-shaped) crack of Sack (cf. §9).

§11. EXPERIMENTAL VERIFICATION OF THE GRIFFITH THEORY

According to the Griffith theory, the observed tensile strength is lower than the maximum stress that the interatomic forces can resist; it ought to be possible,

therefore, to increase the tensile strength by removing or making ineffective the most dangerous cracks. An impressive proof of this was given by Griffith (1920). On freshly drawn glass rods he measured strengths (in bending) up to 63,000 bars, whereas the usual strength of the material was only 1,830 bars. The high strength, however, fell to the normal value within a few hours. With silica glass, occasionally even somewhat higher values were observed; significantly, its high strength did not disappear spontaneously if the material was pure enough (§12). It disappeared, however, instantaneously when the silica rod was touched with any hard body, for example, another silica rod. The strength in bending remained high if the point at which contact took place was on the compressed side.

Griffith's experiments show that, with glasses, the dangerous cracks are at the surface. This has since been found to be true for most cases of brittle fracture.

Joffé, Kirpitschewa, and Levitzky (1924), Joffé (1928) tried to eliminate surface cracks on NaCl crystals by dissolving the surface in warm water while the crystal was under tension; they claim to have measured in this way strengths up to 16,000 bars, while the normal strength of their material was about 50 bars *.

Schmid and Vaupel (1929) could not entirely confirm this, but they also observed increases of strength up to 25 times. Joffé attributed the effect to the removal of surface cracks; Ewald and Polanyi (1925), on the other hand, thought that the primary effect of the water was to reduce the yield stress. The increased plastic deformation would then give rise to "strain strengthening", an increase of strength due to plastic deformation (the propagation of cracks is hindered by the warpings and structural injuries produced by the deformation). Whichever is its primary cause, the existence of the "Joffé effect" proves that NaCl is also capable of withstanding stresses far above its usual strength.

Soon after the publication of Griffith's papers, the view was put forward by Smekal (1927) that, in the case of crystals, the discrepancy between the molecular cohesion and the observed strength would be a consequence, not of accidental flaws, but of a universal "block structure" of crystalline matter. According to this view, crystals would consist of very small blocks within which the strength would have its theoretical value; the cohesion between the blocks, however, would have the experimentally observed low value of the tensile strength. If this were true, the strength of crystals could not be raised by the elimination of cracks. While the interpretation of the Joffé effect remained somewhat ambiguous, a clear decision was reached in the case of mica. The usual tensile strength of mica is between 2,000 and 3,000 bars; it was plausible to assume that this relatively low value was due to cracks at the edges of the sheets, since the very perfect cleavage plane could hardly contain effective cracks. The edges of a mica tensile specimen could be made stressless by the use of grips narrower than the width of the lamella, so that only the central part of the lamella was under stress, and in this way tensile stresses exceeding 30,000 bars could be measured (Orowan 1933 a). In subsequent years, an overwhelming amount of evidence has accumulated against the block structure hypothesis.

In addition to the experimental verifications of the Griffith theory of brittle strength, many industrial processes dating from earlier times are based on the

* The natural rock-salt crystals they have used were hardened by plastic deformation under earth pressure.

circumstance that the usual limiting factor which determines the brittle strength is the presence of surface cracks. In these processes, the strength is increased by the production of a tangential compressive stress in the surface layer; by this expedient, tensions acting on the surface cracks are relieved and perhaps even the growth of surface cracks by devitrification etc. prevented; the concomitant tensile stress in the interior is harmless, partly because its value is much lower, partly because there are no dangerous cracks inside the body. Thus, toughened glass is produced by quenching the surfaces with an air blast; when the interior cools down, its contraction produces a compressive stress in the already colder surface layers. The tensile stress of porcelain can be increased considerably by the use of a glaze that, owing to a lower coefficient of thermal expansion, suffers tangential compression on cooling. If the coefficient of expansion of a glaze is higher than that of the porcelain base, the strength is reduced by its application.

As mentioned in §2, fatigue fracture has many of the characteristics of brittle fracture; one of the common features is that the fatigue limit of metals can usually be raised by producing a residual compressive stress in the surface. This can be achieved, for instance, by shot-blasting ("peening"), surface rolling, or, with steels, by nitriding which causes an expansion of the surface layer.

§12. LENGTH OF THE GRIFFITH CRACKS; MEANS OF OBSERVATION; MECHANISMS OF CRACK FORMATION

By measuring the surface tension of liquid glass at different temperatures, the surface energy of glass at room temperature can be obtained approximately by extrapolation; from this and the measured values of Young's modulus and of the tensile strength, the depth c of the surface cracks can be calculated by using equation (14) or (14a). The crack depth was found to be about 1 or 2 μ for the glasses investigated by Griffith. Such cracks need not be observable microscopically, because their width is a very small fraction of the wavelength of light. However, Andrade and Tsien (1937) could make surface cracks in glass visible by development in sodium vapour. They found, for instance, that freshly drawn pyrex glass contained only few cracks; their number increased, however, very much when the sodium vapour treatment was applied 50 hours after drawing. Figure 5 shows cracks on the internal surface of a freshly drawn pyrex tube and Figure 6 cracks on a microscopic slide. Figure 7, by Andrade and Martindale (1935), shows silver particles coagulated on heating of a film of silver sputtered on a polished disc of silica glass. The 90° bend could hardly arise if the particles were deposited along ordinary scratches. (For Figures 5, 6, 7 see Plate I, opposite p. 208).

In mica, microscopically observable features of the edge cracks and of their mode of propagation under stress give indications for understanding the differences between the strengths of lamellae with edges cut with different methods (Orowan 1933a).

Griffith's observation that the increased strength of freshly drawn silica rods fell to the normal value on contact with any hard body shows that cracks can be produced mechanically. The spontaneous decrease of the abnormally high strength of freshly drawn ordinary potassium glass, in contrast to the behaviour of silica glass, shows that cracks can also arise by chemical attack on the surface (devitrification). The inferior strength properties of soda glass are probably mainly due to its low resistance to atmospheric attack; old soda glass tubes

crack on heating in the flame, but the tendency to cracking can be removed by washing the tube with dilute hydrofluoric acid.

For many crystals (e.g. NaCl or Zn, both of which can break with a typical cleavage fracture) the Griffith formula (14) would demand impossibly large crack depths, often larger than the dimensions of the crystal. This was regarded as a failure of the Griffith theory until it was realized that the Griffith crack propagation mechanism may not be applicable to ductile materials where, on the other hand, entirely different fracture mechanisms can operate. This problem will be treated in Section IV.

§13. STATISTICS OF CRACKS

An important question raised by the Griffith theory is whether the number of the most effective cracks in the specimen is very large or not. The scatter of brittle strength measurements indicates that the number of the dangerous cracks cannot be very large; another indication is that the mean spacing of the cracks on glasses developed by sodium vapour in the experiments of Andrade and Tsien (1937) was of the order of 0·1 mm. A direct proof for the scarcity of cracks, however, was given by Reinkober (1931) who disrupted silica glass fibres, then the two fragments obtained, then the fragments of the fragments, and so on. Owing to the progressive elimination of the most dangerous cracks, the strength rose from generation to generation of the fragments; in one experiment, a fibre of 12 cm. length broke at 2,750 bars, while one of its fragments of 0·6 cm. length had a strength of 7,600 bars.

Reinkober's experiments show that, in general, the brittle strength of a specimen must decrease with its size, because the probability of the presence of an unusually deep crack decreases with the size. Much interesting work on the statistics of brittle strength has been done by a number of authors, among whom Peirce (1926), Weibull (1939a, 1939b), Kontorova (1940), Frenkel and Kontorova (1943), Fisher and Hollomon (1947), and Epstein (1947) should be mentioned. The space available for the present Report does not admit of even the briefest review of the results obtained; however, a few typical difficulties, not all of which have received due attention in the literature, must be mentioned. To begin with, one must make a more or less arbitrary assumption about the distribution function of the crack lengths; some of the simplest functions, however, may lead to senseless results if the number of the cracks is very small or very large (thus, the Gaussian distribution used by Frenkel and Kontorova gives negative strengths if the number of the cracks is very large (see Fisher and Hollomon 1947). Another objectionable point in the existing statistical treatments of brittle strength is the assumption that the number of the cracks is proportional to the volume of the specimen. This makes the results inapplicable to the overwhelming majority of brittle materials (e.g. glasses) where the dangerous cracks are on the surface, so that their number cannot be proportional to the volume. With glasses and similar materials, not even the assumption of a proportionality between the number of the cracks and the surface area is likely to be correct: the experiments of Andrade and Tsien (cf. Figure 5) show that the cracks on the internal walls of glass and silica tubes are usually circumferential, and their number is probably proportional to the length of the tube rather than to its surface area. Similar conditions may exist in the case of rods and fibres also. Any statistical treatment of the brittle strength, therefore, must take into account the individual natural

history of the cracks in the material considered; without closest attention to this, the results are likely to be more interesting mathematically than correct physically.

§14. SIZE EFFECT

It was found by Griffith (1920) that the tensile strength of glass fibres increases rapidly with decreasing thickness; the dependence of the strength σ upon the diameter d could be expressed approximately by the formula

$$\sigma = A - B/d \qquad \ldots\ldots(16)$$

found by Karmarsch in 1858 to represent the strength of thin metal wires as a function of the diameter; A is the strength of large specimens, and B a constant. Reinkober's experiments with thin silica fibres conformed with the Karmarsch formula even better than did Griffith's measurements; a similar behaviour was found with sodium chloride (Jenckel 1932), and a few other materials. (It may be mentioned that Karmarsch's results on metals were not confirmed; they were probably due to the greater amount of cold work received by thinner wires in drawing.)

The cause of the size effect is not known with certainty; many possible explanations have been suggested. An increase of the strength with decreasing thickness would follow from statistical considerations; as was mentioned in the preceding section, however, the existing statistical treatments do not apply to the probable conditions in fibres where the number of the cracks is likely to be proportional to the length rather than to the surface area of the volume. In any case, a statistical effect could hardly explain the Griffith–Karmarsch curve. The point has been made (Orowan 1933 b) that the strength begins to rise rapidly when the diameter of the fibre approaches the order of magnitude of the Griffith crack length for large specimens; the thinnest fibres that can be investigated have diameters less than the ordinary Griffith crack length. They cannot contain, therefore, cracks as long as those on large specimens and, unless their cracks are large enough to weaken the fibre by reducing its cross section, the strength ought to be higher than that of a thick rod. What is known about the origin of cracks also suggests that their depth is likely to be smaller on very thin specimens.

Griffith thought that the increased length of thin fibres of glass was due to the alignment of long molecules during drawing which would reach a higher degree in thin threads: this idea, however, is not supported by the present knowledge of the constitution of glasses. A particularly interesting relevant observation has been reported by Reinkober (1932); he found that both Young's modulus and the shear modulus as determined from torsion tests depend on the thickness of silica fibres in much the same way as the tensile strength. This effect would deserve further investigation; if it is real, the dependence of Young's modulus supports Griffith's hypothesis, while that of the shear modulus is rather unexpected.

§15. ADSORPTION EFFECTS

The strength of mica (Florenski 1930) and of silica fibres (Schurkow 1932) depends on the medium in which the fracture takes place; silica fibres baked out and broken *in vacuo* are 3·5 to 4·5 times stronger than those tested in air. This is obviously due to the fact (Orowan 1933 b) that adsorption from the atmosphere (probably of moisture) diminishes the surface energy and thus, according to the Griffith formula (14), reduces the strength. The surface energies of

glasses in air and *in vacuo* are not known; they could be measured, however, for a material not unlike chemically, i.e. mica (Obreimoff 1930). The surface energy of mica is about 375 erg/cm² in air, and it rises to 4,500 erg/cm² *in vacuo* (Orowan 1933b); if the same ratio is assumed for silica glass, its vacuum strength should be $\sqrt{(4{,}500/375)} \sim 3{\cdot}5$ times higher than the strength in air, in good agreement with the experiments.

§16. THE STATIC FATIGUE OF GLASS

Glasses sometimes break spontaneously; the fracture must be due to internal stresses, and the problem is to explain how a delay of months or years between the manufacture and the fracture can arise. Experimental investigations into this question have shown that the breaking stress of glasses decreases with increasing duration of loading; if the load acts for weeks or months, fracture occurs under stresses that are between a third and a quarter of the strength measured in short-time experiments (Holland and Turner 1940, Preston 1942).

The delay cannot be explained by the assumption of a thermal activation mechanism for the crack propagation. If the Griffith condition is not fulfilled, any propagation of the crack would mean the isothermal creation of free (surface) energy from thermal energy, in contradiction to the Second Law; once the condition is satisfied, however, there are no appreciable energy barriers to be overcome during the cleavage process because the tip of the crack contains numerous molecular bonds in all stages of separation, and the energy barriers for breaking the individual bonds average out.

The explanation of the delayed fracture which has subsequently been confirmed by experiments is that the surface energy of a surface crack is lowered by atmospheric adsorption, and thus the crack propagating stress is reduced if the propagation is slow enough for the adsorbed material to diffuse to the tip of the crack during its propagation (Orowan 1944). The adsorption may be a physical or a chemical one; it may amount to a surface devitrification, and its effect may be reinforced by the mechanical action of the devitrification products which may occupy a greater volume than the glass and so force apart the crack. The ratio of the strengths for very fast and very slow loading, according to the preceding section, must be the square root of the ratio of the surface energies without and with adsorption; if, again, this ratio is assumed to be the same as with mica, the strength in fast loading should be $\sqrt{12} \simeq 3{\cdot}5$ times higher than in prolonged loading, in good agreement with experiments. Furthermore, the long-time strength in experiments carried out *in vacuo* should be as high as the short-time strength; this conclusion has also been confirmed experimentally (Baker and Preston 1946).

§17. BRITTLE FRACTURE UNDER BI- AND TRIAXIAL STRESS

A fracture condition for biaxial states of stress has been derived by Griffith (1924) directly from the Inglis solution of the stress distribution around an elliptical hole in a plate. His assumption is that the (isotropic) material contains cracks of all directions and that the maximum crack length, as well as the radius of curvature at the tip of the crack, is the same for cracks of all orientations. The Inglis solution then gives directly the maximum tensile stress at the tip of the crack of the most dangerous orientation relative to the principal stress axes, as a function of the two principal stresses P and Q. Griffith assumes that

fracture occurs when the highest local tensile stress at the longest crack of the most dangerous orientation reaches a fixed critical value equal to the molecular cohesion (theoretical strength) of the material; those combinations of P and Q at which fracture takes place can be expressed then by equations in which the only physical constant is the ordinary tensile strength K for uniaxial stressing. These equations represent the fracture criterion for biaxial stressing. If tensile stresses are regarded as positive and $P>Q$, the fracture criterion is as follows:

$$\text{If } 3P+Q>0, \text{ fracture occurs when } P=K. \quad \ldots\ldots (17)$$

$$\text{If } 3P+Q<0, \text{ fracture occurs when } (P-Q)^2+8K(P+Q)=0. \quad \ldots\ldots (18)$$

The graphical interpretation of these equations is given in Figure 8 (Orowan 1949b). If P and Q are plotted as rectangular coordinates, equation (18) is

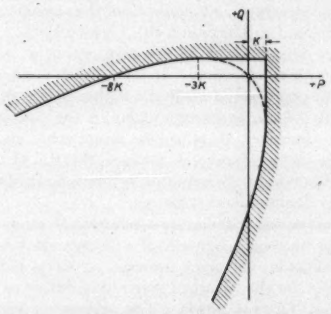


Figure 8. Geometrical representation of the biaxial fracture criterion of Griffith.

that of a parabola which shows its concave side in the direction of the bisector of the negative P and Q axes; equation (17) is that of the vertical tangent of this parabola. If all combinations of P and Q are considered, not only those for which $P>Q$, the fracture conditions have to be completed by $Q=K$ for $3Q+P>0$ and $Q>P$; this means the horizontal tangent to the same parabola. The resulting fracture curve is that drawn with full lines in Figure 8; fracture occurs when the point representing the state of stress crosses the curve in the direction towards its shaded side. An interesting feature is that, on this criterion, a brittle material can withstand any shear stress provided that a sufficiently high hydrostatic stress is superposed*. Equation (18) shows directly that the hydrostatic pressure needed to prevent fracture at the shear stress $\tau=(P-Q)/2$ is $\tau^2/4K$. For uniaxial compression ($P=0$, $Q<0$), equation (18) gives the strength as $8K$, that is, the compressive strength in brittle fracture should be eight times higher than the tensile strength. This relationship is approximately fulfilled for cast iron and for many stones and rocks. There is one difficulty, however: according to Griffith's considerations, the most dangerous cracks in uniaxial compression would be those at 45° to the stress, and they should propagate in their own planes.

* This explains why many materials that are brittle in the ordinary tension and compression tests become ductile if a sufficiently high hydrostatic pressure is superposed (Kick 1892, v. Kármán 1911).

Although fracture surfaces at about 45° are common in compressive tests, failure by cracking *parallel* to the direction of compression is almost equally often observed with glasses and stones; sometimes the surfaces of fracture run at 45° if the compression plates are dry but become parallel to the compressive stress if these are lubricated. The problem of the cracking parallel to the stress in compression has not been solved; a few attempted explanations (Fridman 1946) have been unsatisfactory.

The Griffith biaxial criterion can be applied to triaxial states of stress since, in the approximation used, normal or shear stresses in the plane perpendicular to the edge of the crack cannot exert an appreciable influence.

Until lately, the only experimental support for the fracture conditions (17) and (18) was that they gave good approximate values for the ratio of the compressive to the tensile strength for a number of brittle materials. Recently, however, the present writer has found (Orowan 1949 b) that Bridgman's measurements (1947 b) of fracture stresses of brittle materials under superposed high hydrostatic pressures can be correlated satisfactorily with the uniaxial tensile stress by means of (17) and (18). Table 1 shows the hydrostatic pressure p, the superposed tension S at fracture, the principal stresses* $P=S-p$ and $Q=-p$, the

Table 1

Stresses in 1,000 kg/cm²

Material	p	S	$P=S-p$	$Q=-p$	Eqn. No.†	K	K_{exp}
Pyrex	28	15·5	−12·5	−28	(18)	0·71	
Pyrex fire-polished	27	24·5	− 2·5	−27	(18)	2·54	
Pyrex	28	−47	−28	−75	(18)	2·68	
Carboloy	27	58	+31	−27	(17)	31	between 20 and 30
Carboloy	27	54·5	+27·5	−27	(17)	27·5	between 20 and 30
Beryllium	27·5	18·5	− 9	−27·5	(18)	1·17	<2
Beryllium ‡	18	8	−10	−18	(18)	0·3	<2

† Equation giving fracture condition.

‡ The second measurement with Be was made without a sheath protecting the specimen from the pressure liquid, so that this could penetrate into the cracks. In calculating the value of S for beryllium, both the reduction of the cross-sectional area due to necking, and the inhomogeneity of the stress distribution were taken into account as accurately as was possible with the experimental data available. For calculating the maximum axial tension from the mean tension, Bridgman's formula (1944) was used.

number of the equation that represents the fracture condition for the case in question, the value K calculated from the fracture condition (17) or (18), and the probable value K_{exp} of the fracture stress in simple uniaxial tension as indicated by Bridgman.§ The agreement is as good as can be expected. In particular, the fracture stresses of pyrex glass under tension and under compression with a hydrostatic pressure of 28,000 kg/cm² superposed lead to values in good agreement between themselves if the fire-polished specimen is considered, and with the value to be expected for the ordinary tensile strength.

* For the compressive test (as Pyrex: third row in the Table) the algebraically greater principal stress is $P=-p$, and the smaller principal stress $Q=S-p$.

§ The Table contains only those of Bridgman's measurements which were not obviously unsuited for evaluation. Omitted are those materials which suffered large plastic deformation before fracture (NaCl and phosphor-bronze), those which contained gross defects (samples of Al_2O_3), and a porous material which was crushed by the hydrostatic pressure (pipestone).

§18. BRITTLE STRENGTH OF POLYCRYSTALLINE AGGREGATES

Experience shows that the strength increases with decreasing grain size and that, moreover, the brittle fracture of polycrystalline metals is a coherent crack propagation process: it begins in one grain and spreads to others in the surroundings. The first question is whether the initiation of a crack in a grain, or its propagation to the next grain, needs a higher stress. There is no certain answer to this question, but apparently the propagation of a crack across grain boundaries is, in general, more difficult than its initiation. There are two possibilities for the extension of a crack from one grain to another. Either it may directly penetrate the grain boundary; or it may produce in the neighbouring rain a high stress which starts there an independent crack*. In either case, the condition for the continuation of the propagation across the boundary is probably the attainment of a critical stress at the end of the crack (i.e. at the boundary), or within a certain region in the next grain. Since the stress at a given distance from the tip of the crack is proportional to the square root of its length (cf. equation (11) and Inglis (1913)), one would expect the brittle strength of a polycrystalline metal to be inversely proportional to the square root of the mean grain diameter (Orowan 1933b). This was confirmed in the case of polycrystalline zinc, and recently the same relationship was found between the grain size and the brittle strength in several steels; the brittle strength was measured by fracture in liquid nitrogen (Petch 1949).

IV. DUCTILE FRACTURES AND NOTCH BRITTLENESS

Non-elastic deformations play an essential rôle in many types of fracture; the term "ductile fracture", however, will be used only for those non-brittle fractures in which plastic slip, mechanical twinning, or kinking (Orowan 1942)—i.e. intracrystalline processes of plastic deformation—play a decisive part. Although fractures by mechanical fatigue may also fall within this definition, they will be treated separately.

§19. RUPTURE BY NECKING AND BY SLIPPING-OFF

As already mentioned (§2), rupture cannot be regarded as a fracture in the narrower sense of the word. It is a consequence of an instability of plastic deformation and occurs when the deformation becomes localized in a small part of the specimen. In the case of tensile specimens, the nature of this instability has been discussed in Chapter II. As a rule, necking is followed by fracture in the neck; sometimes, however, the neck becomes progressively thinner and separation occurs without any observable fracture. This process of rupture is promoted by low values of the transverse tension in the neck (see §21); consequently, it occurs much more frequently in sheets than in rods, and it can be provoked by the application of a hydrostatic pressure which counteracts the transverse tension.

Rupture by slipping-off occurs with single crystals, particularly those with one strongly preferred set of slip planes (e.g. Zn, Cd, or Mg). The underlying instability of the slip process is the concentration of the deformation in single slip planes or in clusters of them. This concentration is promoted by slow deformation and high temperatures.

* Private communication from Professor N. F. Mott.

§20. GENERAL FEATURES OF DUCTILE FRACTURE

The only work required for brittle fracture is that needed to overcome the cohesive forces between the atoms at either side of the cleavage crack. This work is so small that, in almost all practical cases, it can be supplied from the elastic energy stored in the specimen and the testing machine at the moment of fracture. Since, in addition, cleavage does not require thermal activation, the velocity of crack propagation in brittle fracture can be very high, of the order of the velocity of sound. In ductile fracture, on the other hand, the work done against the cohesive forces is usually negligible compared with the work of plastic deformation that has to be done in order to extend the crack. The elastic energy in a conventional tensile or torsion specimen is quite inadequate to supply the necessary plastic work, and so the crack can spread only if the external forces continue to do work; its propagation can be arrested at any time simply by stopping the drive of the testing machine.*

The slow propagation of the crack in ductile fracture is shown by the tensile load-extension curve Figure 9. The curves usually seen in the literature end

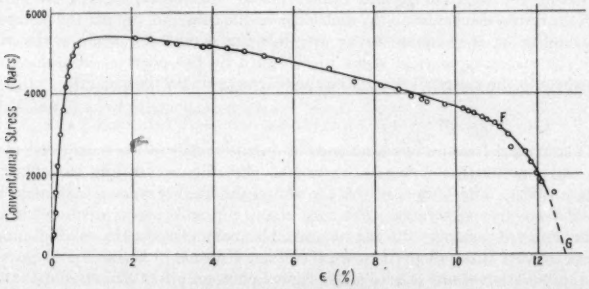


Figure 9. Stress–strain curve of cupro-nickel (1/3 Ni), showing the part corresponding to the propagation of the crack (F to G).

at the point F where the crack propagation begins; however, if a sufficiently rigid testing machine is used, the part FG representing the spread of the crack across the specimen with a gradual decrease of the cross-sectional area to zero can be recorded (Orowan, Nye and Cairns 1945). With $\frac{1}{2}$ in. diameter tensile specimens of pure Cu or Al, the extension of the specimen between the points F and G is of the order of 1 mm.

Obviously, the Griffith theory cannot be applied to ductile fracture where the cohesive work and the change of the elastic energy during crack propagation are insignificant compared with the work done by the external forces, most of which is used for the plastic deformation necessary for the propagation of the crack.† Disregard of this fundamental point has led to much confusion in the physical interpretation of ductile fracture.

* Provided that the dynamometer of the testing machine is sufficiently rigid. The beam, hydraulic piston, or spring dynamometers of the commercial testing machines "run away" as soon as the load begins to drop; such machines are unsuitable for studies of the ductile fracture process.

† *Note added in proof.* A recent development has shown that the Griffith formula can be generalized so as to apply to "brittle" fracture in low carbon steels, where the plastic deformation is concentrated in a thin layer at the surface of fracture; see footnote on p. 214.

As mentioned already, ductile and brittle fractures differ not by the presence or absence, respectively, of plastic deformation before fracture, but by their different mechanisms of crack propagation; typical brittle fracture can occur after considerable plastic deformation. This raises an interesting question. If plastic yielding precedes brittle fracture, the yield stress and the brittle strength must be equal at the moment of fracture. According to Griffith's conceptions, however, the true microscopic local stress at the tip of the propagating crack is very much (perhaps 100 or 1,000 times) higher than the observed brittle strength; how can such high tensile stresses (which must be accompanied by shear stresses of the same order of magnitude) exist in a material the yield stress of which is 100 or 1,000 times smaller? The answer is that shear stresses very much higher than the macroscopic yield stress can exist in very small regions without causing plastic yielding (Bragg 1942); it has been shown that the molecular shear cohesion of a crystal lattice must be much higher than its observed yield stress in shear (Frenkel 1926, Orowan 1940) although, in the opinion of the present writer, the traditional method of estimating the molecular shear cohesion is faulty and the results far too high (Bragg and Lomer 1949). Obviously, a crack can spread with the brittle mechanism if the molecular tensile cohesion, but not the molecular shear cohesion, is overcome at its tip; this can occur even while at the same time the macroscopic yield stress (determined by the presence of dislocations elsewhere in the material) is equal to the macroscopic brittle strength.

§21. THE TENSILE FRACTURE OF DUCTILE METALS

The typical fracture of round rods of ductile metals in the tensile test is the cup-and-cone fracture shown in Figure 10. (For Figures 10 and 11 see Plate II, after p. 000.) The bottom of the cup shows the fibrous type of fracture, while its sides are shear fractures. Fracture of the cup-and-cone type always begins in the centre of the neck; this can be proved by making axial sections of specimens before fracture is complete (between the points F and G in Figure 9). Figure 11 is a reproduction of one of Ludwik's original photographs (1926) obtained in this way; it shows the crack in the neck of an aluminium tensile specimen. Since the fracture begins with the fibrous portion, its fracture criterion must be that of fibrous fracture.

The fact that the cup-and-cone fracture begins in the centre of the neck is usually regarded as a consequence of the circumstance that the axial tension is highest there. This can be recognized qualitatively in the following way. Since the shear stress vanishes at a free surface, the outlines of the necked specimen shown in Figure 12 must be stress trajectories. For reasons of symmetry the axis of the specimen is also a trajectory, and the other trajectories must run in the way indicated in the figure. Let now a lens-shaped volume be delimited by two trajectories A–A, B–B, of the perpendicular set, symmetrical with respect of the centre of the neck. The axial set of trajectories gives the directions of the forces acting upon the volume AABB; except on the axis, these forces have everywhere a radial component which creates a transverse (radial) tension in the neck. This vanishes at the free surfaces AB and rises to a maximum in the centre of the neck. Let now Y_n be the axial and R the radial tension in a point in the narrowest cross section of the neck; Y_n and R are principal stresses and it can be shown that the third (circumferential) principal stress must be approximately R. The state of tension with the principal stresses Y_n, R, R is equivalent

to a uniaxial tension ($Y_n - R$) with a hydrostatic tension R superposed; since the hydrostatic component has negligible influence upon plastic yielding, the

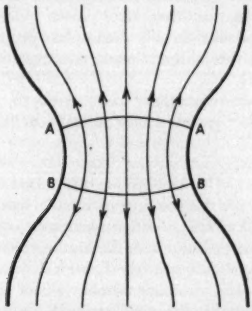


Figure 12. Sketch of the stress trajectories in the neck of a tensile specimen.

uniaxial component must be equal to the yield stress Y in simple tension if the neck is in a state of plastic deformation:

$$Y_n - R = Y \quad \text{or} \quad Y_n = Y + R. \qquad \ldots\ldots(19)$$

Bridgman has shown that the strain is fairly constant over the cross section of the neck (Bridgman 1947a); hence, the yield stress Y is also approximately constant, and the axial tension Y_n rises with R to a maximum in the centre of the neck. It is usually thought that this maximum of the axial tension and of the hydrostatic tension is responsible for the starting of the fracture in the centre. However, the matter is by no means clear since, as will be seen in § 31, fibrous fracture certainly does not obey a maximum tensile stress condition.

An additional reason why the crack starts in the centre is that the central parts of a rolled rod, having arisen from the central parts of the ingot where the segregations are strongest, are weaker than the peripheral parts.

§ 22. THE "TRUE" TENSILE STRENGTH

The "true", as distinguished from the ultimate, tensile strength is defined as the load at which fracture begins, divided by the area of the cross section in which the fracture begins, that is, by the area of the narrowest cross section of the neck in the case of the cup-and-cone fracture. It should be pointed out in this section that this definition does not give the true axial tension at which the fracture begins, and it will be seen in § 31 that the fibrous fracture does not obey a tensile stress condition; thus one cannot speak of a true strength for fibrous fracture in any case.

From the preceding section it is clear that the stress distribution in the narrowest cross section of the neck is not uniform; the "true" strength according to the usual definition, therefore, is not the true stress at which fracture begins but merely the mean stress in the cross section. The problem of the stress distribution in the neck has not been solved yet; approximate solutions by

Siebel (1925) and Bridgman (1944) indicate, however, that the tension in the centre can well be 30 or 40% higher than the mean tensile stress.

If the area of the narrowest cross section of the neck is measured, as usual, on the fractured specimen, another error arises (McAdam and Mebs 1943). Since the neck contracts considerably during the propagation of the crack, its smallest cross section at the moment when fracture starts is greater than after completion of the fracture.

A brief reflection shows that the "true" strength, even if it existed and if it could be measured correctly, would have no practical importance for applications in engineering.

§23. THE LUDWIK HYPOTHESIS

Until recently, work on the fracture of ductile metals was dominated by a hypothesis due to Ludwik (1923). Although it is clear now that this hypothesis may be true only for the brittle but not for the fibrous type of fracture, its discussion is necessary for the understanding of the historical development of the subject.

Ludwik assumed that the mechanical behaviour of a ductile metal was determined by two basic functions, the yield stress Y and the strength S, in their dependence upon the plastic strain ϵ. He further assumed that, in general, the yield stress would increase with the plastic strain more rapidly than the strength; although, for a ductile metal, initially $Y < S$, plastic deformation would raise Y ultimately to the value of S, and then fracture would occur at the point of intersection F of the (Y, ϵ) curve with the (S, ϵ) curve (Figure 13).

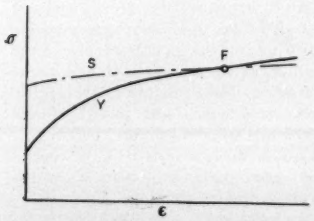


Figure 13. Ludwik's conception of fracture occurring at the point of intersection of a yield stress–strain curve with a strength–strain curve.

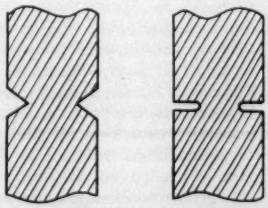


Figure 14. Axial sections of tensile specimens with (*a*) a circumferential V-notch and (*b*) sawcut notch.

When Ludwik put forward the view represented by Figure 13, it was not realized that several essentially different physical processes may lead to the common result of fracture; in fact, his aim was to explain, with the assumption of only one fracture condition, the occurrence of apparently brittle fractures in carbon steels that are highly ductile in the usual tensile test, i.e., the phenomenon called notch brittleness. The exploration of notch brittleness has been a pivotal point in the development of the physics of ductile fracture; before the latter can be treated in greater detail, a review of the main facts and theories of notch brittleness must be given.

§24. NOTCH BRITTLENESS

Ferritic steels sometimes break with an almost completely brittle fracture, while the same material shows a high ductility and fibrous fracture in the ordinary tensile test. Impact loading and low temperatures promote brittle fracture; severe continental winters have often caused numerous brittle fractures of railway rails and rolling stock. The last spectacular epidemic of such failures occurred in welded ships during the 1939–45 war.

A strong tendency to brittle fracture on the part of a steel can be detected by bending tests on notched bars (hence the name "notch" brittleness). In a crude form, such tests were used by blacksmiths hundreds of years ago; the more refined modern form of the test (breaking the notched bar specimen by the impact of a pendulum hammer, the energy loss of which measures the work of fracture) was introduced by S. B. Russell (1897). The conventional specimens and testing machines used at present were designed by Charpy and by Izod.

Not so long ago, brittle fracture in steels was attributed to high speeds of (macroscopic) deformation, and this belief has not yet completely disappeared. However, slow bending is usually almost as effective as impact bending in producing brittle fracture; thus, although high rates of loading increase the tendency to brittleness, they cannot be its main cause. It was Ludwik who drew attention to what is at present regarded as the main direct cause of notch brittle behaviour: the triaxial state of tension arising on plastic deformation in specimens containing notches or cracks. It was shown in §21 that there is a transverse (radial) tension in the neck of a tensile specimen. Obviously, the same argument applies to a rod-shaped tensile specimen provided with a circumferential notch (Figures 14(*a*), 14(*b*)): as soon as plastic yielding has extended over the notch core (the volume surrounded by the circumferential notch) and the difference between the axial and the radial tension has become equal to the yield stress Y (equation (19)), the axial stress Y_n rises from the value Y of the yield stress at the tip of the notch to a maximum in the centre of the core, the value of which increases with the depth and sharpness of the notch.*

In the presence of a notch, therefore, the highest tensile stress in the specimen is not the yield stress Y but the axial stress $Y_n = Y + R$ (see §21) in the centre of the notch core. The curve of the greatest tensile stress as a function of the plastic strain is then obtained from the yield stress–strain curve $Y(\epsilon)$ by addition of the transverse stress R (Figure 15); with the strength–strain curve assumed in this figure, the point of fracture moves from F to F_n as a consequence of the circumferential notch, and the strain at fracture diminishes from that corresponding to the abscissa of F to that of F_n; that is, the notch reduces the deformation of the specimen. This was Ludwik's theory of notch brittleness.

A few years after Ludwik's publication, Fettweis (1929) pointed out in a valuable monograph on notch brittleness that Ludwik's theory did not really explain the fundamental physical difference between the brittle fracture with its bright crystalline cleavage faces and the ductile fracture with the dull velvety surface of fracture, between which there is no continuous transition; in borderline

* Before plastic flow has set in, the tensile stress in the notched cross section is highest at the tip of the notch and falls towards the centre of the section. Plastic deformation begins at the tip of the notch and keeps the stress there approximately constant while the stresses in the elastic central part of the section are rising. Finally, a stress maximum develops in the centre as plastic deformation extends over the notch core. It is easily seen that a triaxial state of tension arises around any notch or crack (not only a circumferential one) if the applied tension has a component perpendicular to it.

cases between brittle and ductile fracture, a part of the surface of fracture is purely crystalline, and the rest purely fibrous. The missing element in the

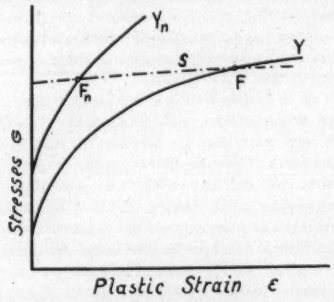


Figure 15. Ludwik's theory of notch brittleness: Y_n is the axial yield tension raised by the triaxial tension due to a notch.

theory was supplied by Davidenkov (1936) who suggested that, corresponding to the two types of fracture observed, there would be two strength–strain curves: one (B in Figure 16) for brittle, and another, D, for ductile fracture. He assumed

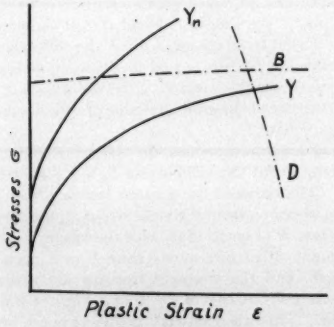


Figure 16. Davidenkov's modification of the Ludwik theory: D is the hypothetical ductile strength, B the brittle strength.

the positions of these curves as shown in Figure 16; the reason for assuming the curious downward slope of the ductile strength curve will be explained in § 30. In the absence of a notch or neck, the highest tensile stress as a function of the plastic strain is given by the yield stress curve $Y(\epsilon)$, and ductile fracture would take place at its intersection with the ductile strength curve D. With a notch or a crack, the curve of the highest axial tension $Y_n = Y + R$ might intersect the curve B first, and then brittle fracture would occur at a greatly reduced value of the fracture strain.

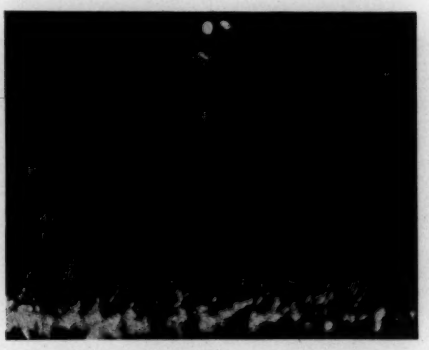


Figure 5. Cracks developed by Na vapour on internal surface of a Pyrex tube; they are perpendicular to the tube axis.—Andrade and Tsien (1937).

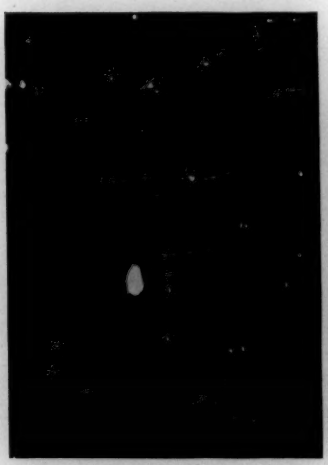


Figure 6. Cracks developed by Na vapour on a soda glass plate.

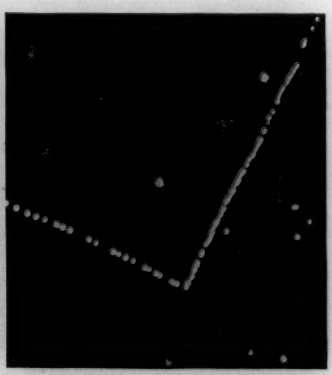


Figure 7. Cracks revealed by the coagulation of cathode-sputtered Ag on a silica glass disc.—Andrade and Martindale (1935).

Plate I.

To face page 208

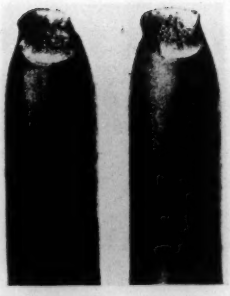


Figure 10. Cup-and-cone fracture of mild steel. About half of the cup and of the cone is on each fragment.

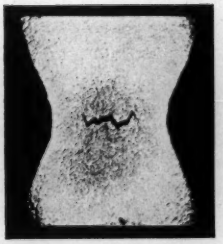
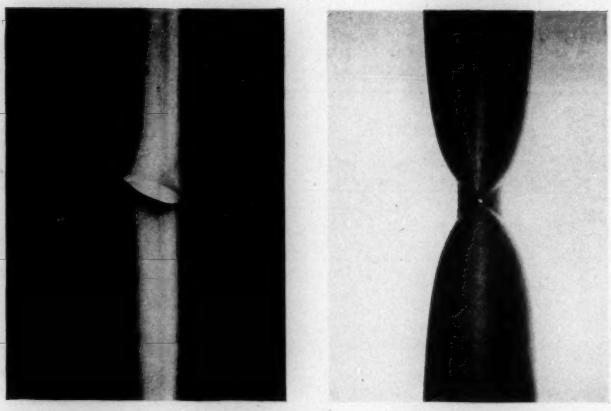


Figure 11. Crack in the neck of an Al tensile specimen, shown by sectioning before completed fracture. — Ludwik (1926).

Plate II.



(*a*) (*b*)

Figure 25. Channel going through the neck of an extended Al crystal.—Mrs. Tipper (*née* Elam); photo Mr. S. D. Charter. Magnification 2×.

Figure 21. Electron micrograph of intercrystalline temper brittle fracture of a steel.—Nutting (1949). Alumina film replica from Al pressing ; magnification about 500×.

Figure 35. Fatigue fracture of a steel roll from a cold strip mill

Plate III.

As already mentioned, it will be shown in §31 that fibrous fracture does not occur at definite values (which may depend on the strain) of the highest principal tension; in other words, with different triaxialities of tension and different forms of the specimen, the points in the stress–strain diagram at which fibrous fracture occurs do not lie on a curve. Consequently, Davidenkov's scheme must be modified by omitting the (non-existing) ductile strength curve D (Orowan, Nye and Cairns 1945, Orowan 1945). The final diagram representing the condition for brittle fracture is then that of Figure 17. In the absence of a sharp

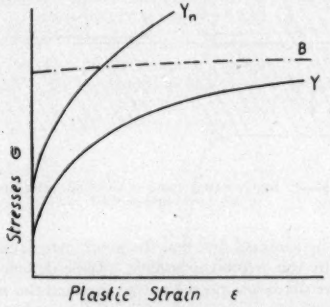


Figure 17. Present version of the Ludwik theory: there is no curve of ductile fractures.

notch or crack, the Y_n curves end by ductile fracture in points the locus of which is not a curve. With a sufficiently sharp notch or crack, the Y_n curve may intersect the curve B and then brittle fracture occurs.

§25. CRITERION FOR FULLY DUCTILE, NOTCH BRITTLE AND BRITTLE BEHAVIOUR

The triaxial tension in the core of a plastically deformed circumferential notch can be regarded as being due to the constraint imposed upon the core by the larger cross sections above and below which hinder the transverse contraction and thus the axial extension of the core. In this sense, the ratio of the axial yield tension Y_n to the uniaxial yield stress Y can be called a "plastic constraint factor"; the mean axial tension in the narrowest section of the core, divided by Y, is then the *mean constraint factor*, and the highest value of Y_n, in the centre of the core, divided by Y, the *maximum constraint factor*.

Until fairly recently, it was believed (see §30) that the constraint factors increase to infinity if the core ratio $= \dfrac{\text{minimum notch core area}}{\text{area of the unnotched cross sections}}$, as well as the radius of curvature at the tip of the notch, and its width become vanishingly small ("ideally deep and sharp notch"). If this were so, a sufficiently deep and sharp crack could induce brittle fracture in any material, since it could raise the tensile stress to any high value. This is in disagreement with the fact that genuine brittle fracture cannot be observed in most of the very ductile metals and alloys. The reason for this is that, in reality, the maximum constraint

factor of an ideally deep and sharp notch, far from being infinity is only of the order 3. This can be proved in the following way (Orowan, Nye and Cairns 1945, Orowan 1945).

Figure 18 (*a*) shows an ideally deep and sharp notch in an infinitely extended

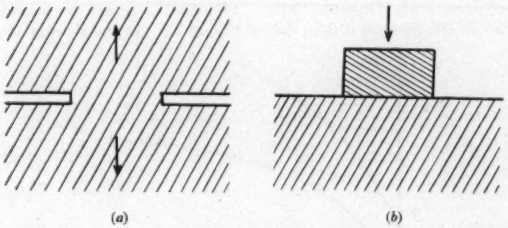


Figure 18. (*a*) ideally deep and sharp notch in an infinite solid, and (*b*) flat indenter on a semi-infinite solid.

plastic solid; let it be assumed first that the notch core is circular. If a vertical tension (indicated by the arrows) is applied, plastic deformation takes place in the two semi-infinite solids around the notch core, and the stress distribution in either of them will differ from that in a semi-infinite solid indented by a frictionless circular punch (Figure 18 (*b*)) only in that the signs of all stress components are reversed. According to an approximate calculation of the punch indentation problem by Hencky (1923), for an ideally plastic solid the highest value of the axial stress (which occurs in the centre of the indentation) is about $3{\cdot}3\,Y$; the mean pressure over the contact area is about $2{\cdot}9\,Y$. Hence, the maximum constraint factor of an ideally sharp and deep notch in an ideally plastic material is about 3·3, and the mean constraint factor about 2·9.

In considering the corresponding two-dimensional problem, Figure 18 (*a*) must be regarded as a section of two semi-infinite solids welded together along an infinitely long strip representing a notch core of infinite length perpendicular to the plane of the figure. This case is perhaps more appropriate in considering the effect of more or less straight cracks in a large bulk of metal. The plastic stress distribution, apart from the signs of the stress components, is the same as in a semi-infinite plastic solid indented with a frictionless indenter along an infinitely long narrow strip. This stress distribution problem has been solved for the ideally plastic solid by Prandtl (1923) (see also Hill 1949). The stress is constant over the area of contact and is equal to $(1-\pi/2)Y \simeq 2{\cdot}57\,Y$. The maximum, as well as the mean, constraint factor for the two-dimensional ideal notch in an ideally plastic solid, therefore, is 2·57.

Since no crack or notch can, by its plastic constraint, raise the tensile stress above $2{\cdot}57\,Y$ or (if the axially symmetrical case is more fitting) $3{\cdot}3Y$, a material can show notch brittleness only if its brittle strength, at a stage of plastic deformation, becomes less than about $2{\cdot}57Y$ or $3{\cdot}3Y$. If the brittle strength is always above $3{\cdot}3Y$, no brittle fracture can take place. This seems to be the case with most very ductile metals,

If the alternative constraint factors 2·57 and 3·3 are replaced approximately by 3, the ductility properties of metals can be said to depend on the relative magnitudes of the yield stress Y (in tension) and the brittle strength B in the following manner:

if $B<Y$, the material is brittle;
if $Y<B<3Y$, the material is ductile in the tensile test but notch-brittle;
if $B>3Y$, the material is fully ductile (not notch-brittle).

§26. THE TRANSITION TEMPERATURE BETWEEN FULL DUCTILITY AND NOTCH BRITTLENESS

If the work of fracture, measured by the pendulum hammer, is plotted for a given material as a function of the temperature, a curve of the type shown in Figure 19 is obtained. Above a more or less narrow temperature interval the

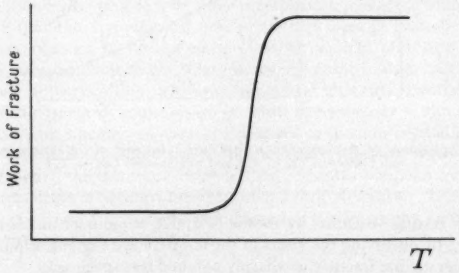


Figure 19. The transition temperature: the "notch impact value" (work of fracture) as a function of the temperature.

work has a fairly constant high value, and the fracture is fibrous; below this interval, the work has a constant low value, and the fracture is of the cleavage type (crystalline). As the temperature is lowered through the transition interval, an increasing fraction of the surface of fracture becomes crystalline.

The existence and the sharpness of the transition are explained by Figure 20, in which B, Y, and $3Y$ are plotted as functions of the temperature. As the Griffith formula suggests, B does not depend strongly on the temperature, because neither the elastic moduli, nor the surface energy change rapidly with T; Y, on the other hand, usually rises by a factor between 1·5 and 3 as the temperature is lowered from 20° C. to that of liquid air. If the curve of B intersects that of $3Y$ at the temperature T_2, the material is ductile above but notch brittle below T_2 while it is completely brittle below T_1, the point in which B and Y intersect. On the simple picture underlying Figure 20, therefore, the transition would occur sharply at a critical temperature. In reality, it takes place within a more or less narrow temperature interval.

It is a much discussed question what types of specimen and test are best suited to give an accurate measure of the tendency to notch brittleness. Figure 20 shows that a measure of this tendency cannot be obtained from measurements at one temperature only; these can show, in general, only whether the transition

interval is above or below the temperature of the measurement. The necessary and usually sufficient data for measuring the tendency to notch brittleness are the position and the width of the transition interval. These data can be obtained

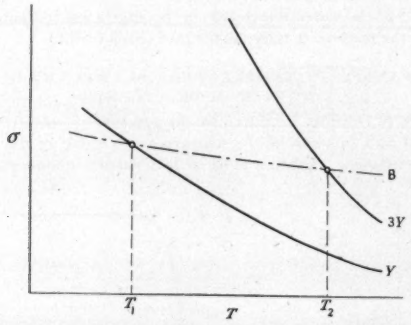


Figure 20. Explanation of the transition temperature. Regions of full brittleness (O to T_1), notch brittleness (T_1 to T_2), and full ductility (above T_2).

without any testing machine, by means of a vice, a hammer and a refrigerator or solid CO_2, by observing the limits of the temperature regions of purely fibrous and purely crystalline fracture on sharply notched test specimens.

§27. NOTCH BRITTLENESS AND LOW TEMPERATURE BRITTLENESS

Cooling down to the temperature of liquid air raises the yield stress of metals roughly 1·5 to 3 times; the presence of notches or cracks, according to §25, also increases the highest tensile stress up to three times for a given value of the yield stress. If it is assumed that the brittle strength varies with the temperature much less than does the yield stress, both notch brittleness and brittleness at the temperature of liquid air must be expected to occur if the brittle strength is less than about two or three times higher than the yield stress at any stage of deformation. The conditions of notch brittleness and low temperature brittleness being approximately identical, metals that are notch brittle at room temperature may be expected to be completely brittle at liquid air temperature. This, in fact, is so with ferritic steels which are almost exceptional among the common metals in combining high ductility in the normal tensile test with notch brittleness and low temperature brittleness.

The direct measurement of the brittle strength of steels at liquid air temperature, after various amounts of deformation at higher temperatures, is one of the most powerful methods of investigating notch brittleness. Sakharov (1936), Hollomon and Zener (1944) have done work on these lines; Petch (1949) has studied the influence of the grain size on the brittle strength by low temperature tensile tests,

§28. TEMPER BRITTLENESS

The ordinary notch brittleness of ferritic steels, particularly plain carbon steels, is a consequence of their relatively low brittle strength for transcrystalline (cleavage) fracture which is just sufficient to make them ductile in the tensile test but not enough to ensure ductility in the presence of a crack or a sharp notch. This seems to be an inherent property of plain carbon steels which cannot be removed radically without the addition of allowing elements, although the transition temperature can be depressed far enough below room temperature by the careful removal of nitrogen, oxygen, phosphorus and sulphur. In addition to this, probably constitutional, type of notch brittleness, steels can develop a more accidental type if, after hardening, they are tempered in a temperature interval that lies usually between 250° and 450° C. The characteristic feature of this "temper brittleness" is that the fracture is prevalently intercrystalline (Figure 21, Plate III) (Nutting 1949); it is due, therefore, to a weakening of the cohesion at the grain boundaries, probably by some precipitation process. Many metallurgists believe that this hypothetical precipitation process is basically related to that which is assumed to be the cause of strain ageing, blue brittleness, and perhaps also of quench ageing. The interesting point is that, as the example of steel shows, polycrystalline metals can undergo at least two essentially different types of brittle fracture, transcrystalline and intergranular; which of these occurs depends on whether the cleavage strength or the intercrystalline cohesion is higher. If a steel is temper brittle, its boundary cohesion is lower than the cleavage strength; the transition interval then separates temperatures of ductile fracture from those of mainly intergranular brittle fracture. Otherwise, the general theory of notch brittleness as indicated in the preceding sections applies to temper brittleness equally well.

§29. SIZE EFFECT IN NOTCH BRITTLENESS

Large specimens of a given material are generally more prone to notch brittleness than small ones; of a series of geometrically similar specimens, with geometrically similar notches, the smallest may be quite ductile in bending while the largest break explosively without any directly visible plastic deformation. This size effect has been the subject of a number of investigations; one of the latest of these is the careful work of Shearin, Ruark, and Trimble (1948) in whose paper references to earlier publications are available.

Once a crack (ductile or brittle) has started at the bottom of the notch, the geometrical similarity between specimens of different size is destroyed, since the curvature at the tip of the crack is determined by the structure of the material, not by its machined shape. This, however, does not explain the size effect; the question is why the first crack, whether ductile or brittle, does not arise simultaneously in a series of geometrically similar specimens. There are only two obvious factors which may lead to a different behaviour according to the size. First, the ratio of the grain size to the specimen size changes with the latter; secondly, the probability of finding more, and more dangerous, natural flaws is higher with larger specimens (cf. the size effect with brittle materials, §14). Shearin, Ruark, and Trimble attribute more importance to the second factor; however, the problem of the size effect cannot be said to be definitely solved.*

* For note see next page.

§30. KUNTZE'S "TECHNICAL COHESIVE STRENGTH"

Starting from Ludwik's hypothesis that fracture is governed by a unique strength–strain curve (§23), Kuntze (1928, 1931, 1932) attempted to develop a general method for obtaining this curve. In recent years, the assumptions on which his method was based have been found to be untenable, and the results obtained devoid of the meaning attributed to them; however, the literature of the subject (Hoyt 1936) is so large and so one-sided that it cannot be ignored in the present review.

It was already recognized by Ludwik and Scheu (1923) that the strength–strain curve S in Figure 13 could be obtained point for point by applying plastic constraints of various magnitudes through providing tensile specimens with circumferential notches of various depths and sharpnesses; in this way different (Y_n, ϵ) curves would be obtained and their end points, representing fracture stresses, would be points of the (S, ϵ) curve. From measurements on notched rods, mainly in the elastic region, Kuntze came to the view that the plastic constraint factor of an ideally deep and sharp notch would be infinitely high; in other words, that a specimen provided with an ideal notch would fracture without any plastic deformation. That this assumption is wrong was seen in §25; the notch constraint factor cannot rise above about three, and at least some plastic deformation is an essential condition for the development of the constraint. Next, Kuntze set up an extrapolation formula to give the fracture stress measured with a standard finite notch; it was shown later (Orowan, Nye and Cairns 1945) that this formula, apart from other objections, was mathematically inconsistent with the manner in which it was used. Kuntze's method of obtaining a point on the strength–strain curve was to measure the maximum load stress of a specimen provided with a standard circumferential V-notch of 60° notch angle, and to multiply it by an extrapolation factor assumed to represent the ratio of the constraint factor of the ideal notch to that of the standard notch used. By repeating this with a number of specimens made of materials which have previously received different amounts of plastic deformation, the strength–strain curve in the sense of Ludwik was obtained. Kuntze's "technical cohesive strength" was simply the strength S plotted in Ludwik's diagram, Figure 3; "technical" meant that it was the observed macroscopic, not the molecular, strength (cf. §8); and the word "cohesive" was to indicate that it was a fracture stress, not a yield stress.

With fairly ductile materials, the plastic maximum load (cf. §3) was reached before fracture occurred in the tests with Kuntze's standard notch; the conventional maximum load stress measured by Kuntze was, therefore, simply the

---

* *Note added in proof.* A possible explanation of the size effect occurred to the writer after the present manuscript was sent to the Editor. A thin layer of some 0·5 mm. depth below the surface of brittle fracture in low carbon steels always shows a few per cent plastic strain. The corresponding plastic work p per unit of area must be added to the surface energy α in Griffith's equation; in fact, the order of magnitude of p (10^6 to 10^7 erg/cm^2) is higher than that of the surface energy (10^5 erg/cm^2), so that the latter can be neglected and, instead of Griffith's equation, the condition

$$\sigma = (pE/c)^{\frac{1}{2}}$$

used; if this is fulfilled, the crack can propagate rapidly from the supply of elastic energy in the specimen. With the observed values of σ and E, the necessary crack length is 1,000 to 10,000 times greater than the ordinary Griffith crack length; it must be of the order of several mm. This may explain why, in ductile steels, brittle fracture is always introduced by a narrow zone of fibrous fracture which creates the crack length demanded by the above condition. Specimens cannot show brittle fracture if their dimensions are below the critical crack length, and full notch brittleness arises only when the crack length of the order of 1 mm. becomes small compared with the dimensions of the specimen. In the transition interval a size effect should be observable.

ultimate stress multiplied by a plastic constraint factor (Orowan, Nye and Cairns 1945). With less ductile materials, fractures took place before the maximum load was reached; such fractures, however, were usually discounted as having occurred prematurely before plastic deformation could level down the elastic stress concentration at the tip of the notch.

As would be expected from the preceding considerations, most of Kuntze's technical cohesive strengths were between 1·8 and 2·8 times higher than the ultimate stress of the material; the main reason why the corresponding plastic constraint factor was usually below 2·8 must have been the circumstance that real metals deviate from ideal plasticity by having a more or less intense strain hardening.

That the constraint factor of the ideal notch is not infinitely high but only of the order of 2 had been concluded by Sachs and Lubahn (1942) from the ratio of the technical cohesive strength to the ultimate stress, before a theoretical calculation of the factor was given. Furthermore, before the real meaning of the technical cohesive strength became sufficiently clear, McAdam and his collaborators (McAdam 1941, 1942, McAdam and Mebs 1943) had already demonstrated experimentally that Kuntze's method could not be correct. Their principal argument was this: A material can be fractured with a given fracture strain in two different ways: either by imparting the entire fracture strain to the (annealed) material before machining the notched specimen, and then applying Kuntze's method which purported to determine the fracture stress without additional deformation; or, by machining the specimen from annealed material and providing it with a notch of such depth and sharpness that the strain would reach the desired final value at the moment of fracture. McAdam showed that the strength–strain curves obtained with these two alternative methods were entirely different. One of his curves is reproduced in Figure 22; K is Kuntze's cohesive strength, and M McAdam's curve obtained with the second of the methods mentioned.

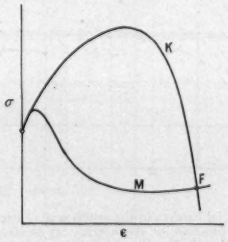


Figure 22. Kuntze's (K) and McAdam's (M) curve of the "technical cohesive strength".

The technical cohesive strength curves obtained by Kuntze showed, in general, the typical shape of curve K in Figure 22; the point F on the sharply descending branch represents the fracture stress of an unnotched specimen. The maximum of the curve and the subsequent descent were interpreted as being due to a breaking-up of the material with increasing plastic distortion (§31). Davidenkov's assumption of a descending ductile strength curve (§24 and Figure 16) arose from

Kuntze's curve which Davidenkov initially interpreted as representing two different curves: he identified the ascending branch with the curve of the brittle strength, the descending one with that of the ductile strength. In reality, there were no measured points on the descending branch, except for the point F and one or two points near the maximum; this branch was obtained by swinging round the curve obtained from notched rod tests up to the maximum or slightly beyond, so that it went through the point F representing the fracture stress of a notchless rod. Since the notched specimens gave ultimate stresses containing a plastic constraint factor while the point F represents a real fracture stress, this point does not belong to the curve whose descending part, therefore, is unreal.

§31. NON-EXISTENCE OF A TENSILE STRESS CRITERION FOR FIBROUS FRACTURE

Figure 22 shows that the "technical cohesive strengths" as obtained by Kuntze's and by McAdam's method differ greatly for the same fracture strain; is it possible that, for a given plastic strain, fracture occurred nevertheless at the same value of the tensile stress? This cannot be decided from Kuntze's and McAdam's measurements since, as mentioned in the preceding section, the measured quantities were usually ultimate stresses under plastic constraint, not fracture stresses. This objection does not apply to Figure 23 in which the

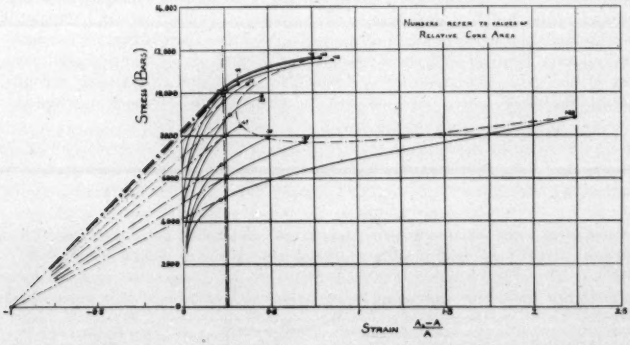


Figure 23. The true mean axial tension in the notch core of copper tensile specimens with thin sawcut notches, as a function of the notch core strain, for notches of different depths (Orowan, Nye and Cairns 1945).

curves drawn with full lines are true mean axial stresses in the narrowest part of the notch core plotted over the true mean linear tensile strain in the notch core, determined from current measurements of the core diameter during the test (Orowan, Nye and Cairns 1945); the notch angle was 0°, that is, the notch was a thin sawcut (of 0·2 mm. width) with a semi-circular tip. The fractional numbers at the ends of the curves are ratios of the notch core area to that of the unnotched cross sections ("relative core area"). The dash-dotted curve is the locus of the

points of fracture. The fracture strain 0·75, for instance, can be obtained either with a relative core area of about 0·8, or with one between 0·01 and 0·03; the true mean fracture stress in the second case, however, is about 50% higher than in the first. Since the axial stress at the notch root must have been equal to the yield stress in both cases, the ratio of the axial tensions in the centre must have been greater than that of the mean tensions; by plotting the true maximum tensile stresses, therefore, the difference between the fracture stresses would increase. Now two possibilities must be considered: either the fracture was of the same type in both cases, or, for instance, it began in the centre in one case and at the notch root in the other. In the first case, the fracture stress could not have been the same for the two types of specimen; in the second, the fracture could not have obeyed a maximum tensile stress condition if it began at the notch root instead of at the centre where the tensile stress is higher. In either case, the fracture could not have obeyed a critical tensile stress condition. Figure 23 shows experiments with mild steel; similar sets of curves leading to the same conclusions were obtained with copper.

That the ductile tensile fracture cannot obey a maximum tensile stress condition can be seen from the fact that the crack opens up wide as it spreads from the centre of the neck outwards (cf. Figure 11 and §20). This requires a large plastic extension in the surroundings of the crack, and thus the fulfilment of the condition of plastic yielding which is substantially a shear stress, not a tensile stress, condition. Even a high axial tension is unable, therefore, to cause ductile fracture if it is opposed by high transverse tensions; these can be produced not only by notching the specimen but, particularly simply and effectively, by reducing its length so that it becomes a thin disc, and then welding this between the flat ends of two round rods of a harder metal (Figure 24). The constraint

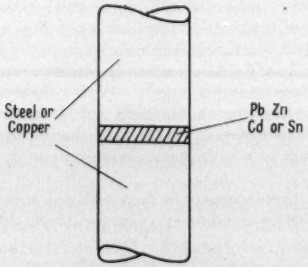


Figure 24. Sandwich tensile specimen.

of the hard rods hinders the transverse contraction, and thus the axial extension, of the softer disc, and so axial tensions considerably exceeding the uniaxial yield stress can be obtained. Such experiments have been carried out by Zschokke and Montandon (1944) and independently by Orowan, Nye and Cairns (1945). In a test made by the latter authors a disc of tin, 0·4 mm. thick and of $\frac{1}{2}$ in. diameter, was welded by pouring between the flat ends of mild steel rods; tensile failure occurred by separation at the boundary layer between the two metals at a mean

axial tension of 1,250 bars*, while the ultimate stress of the cast pure tin used was 225 bars and its "true" strength in the conventional tensile test about 400 bars. Similar results were obtained by Zschokke and Montandon. These experiments prove directly that ductile tensile fracture does not occur at a critical value of the tensile stress.

In connection with the disproof of the normal stress condition for ductile fracture it must be indicated why an apparent experimental proof of the normal stress criterion, put forward by Maier (1935, see also Gensamer 1941), is invalid. Maier carried out tensile tests on thin-walled tubes under combined axial tension and internal pressure; he could produce in this way various combinations of two principal tensions parallel to the wall of the tube. He found that, in all cases, the tension at fracture normal to the surface of fracture was constant (apart from a small anisotropy effect), and this was interpreted as a proof of a normal stress criterion for fracture in general. A closer consideration, however, shows that the observed facts prove nothing of this sort. As soon as necking begins, plastic deformation ceases outside the neck; the neck extends, as it were, between rigid grips. The plastic extension of the neck and thus the propagation of the crack in it depends then only on the tension in the narrowest section of the neck but not on the tension parallel to the trace of the neck along the tube wall; the influence of the biaxiality of the tension is wiped out as soon as necking begins. The shear stress in the cross section of the neck, according to the equations of Lévy and Mises (see Nadai 1931), can produce only a shear strain in this cross section, not the extension normal to it which is necessary for the crack propagation.

§ 32. OTHER SUGGESTED CRITERIA FOR DUCTILE FRACTURE

Fridman (1941, 1946, see also Stanley-Smith 1945) suggested that ductile fractures obey a critical shear stress condition; since shear stress and shear strain are linked up by the yield stress–strain curve, this is equivalent to a maximum shear strain condition of fracture. That such a criterion cannot be correct for the fibrous type of tensile fracture is seen directly from the fact that a metal that breaks in tension after 2- or 3-fold extension in the neck can usually be rolled or drawn to 50 or 100 times its initial length. With a shear stress or shear strain criterion, a hydrostatic pressure should have little influence upon the fracture process; Bridgman's experiments (1947a) with ductile metals, however, have shown that the influence of a hydrostatic pressure upon ductile tensile fracture is very great.

On the other hand, there seems to be no reason why shear fracture should not obey a critical shear strain condition; experiments on this subject, however, are scanty.

§33. DUCTILE FRACTURE AS A CRACK PROPAGATION PROCESS

As mentioned in §30, Kuntze and others interpreted the descending branch of the technical cohesive curve (curve K in Figure 22) as the result of a progressive disintegration of the material as the fracture strain of a notchless specimen was approached; they thought that at the point of fracture the cohesion of the material would rapidly approach zero. That this cannot be so is clear from the fact that ductile metals can be given by rolling or drawing plastic extensions many times larger than the extension in the neck at tensile fracture, without

* 1 bar $= 10^6$ dyne/cm^2.

any loss of strength. Consequently, ductile fracture in tension is not a disintegration of the material but a crack propagation process which leaves the cohesion within the fragments largely unaltered. Another proof of this was given by Swift (1939) who strained rods in torsion by various amounts up to the point at which shear fracture was imminent, and then subjected them to a tensile test. For all torsional strains except those quite near the breaking point there was no sign of any decrease of the "tensile strength".

As was said earlier, the mechanism of the crack propagation in ductile (fibrous and shear) fracture cannot be of the Griffith type; furthermore, the microscopic (local) fracture condition to be satisfied at the tip of the propagating crack cannot be a tensile stress condition because, evidently, the crack extends by means of plastic deformation which obeys either a maximum shear stress, or the Mises–Huber octahedral shear stress, condition. The simplest assumption is that the crack propagates if the plastic strain at its tip reaches a critical magnitude which, of course, may exceed the macroscopic fracture strain considerably, just as the tension at the tip of the Griffith crack exceeds the mean tension in the specimen. The local plastic strain condition is equivalent to a local shear stress condition, since the two quantities are connected by the yield stress–strain curve. In order to obtain on this basis the macroscopic fracture condition corresponding to the Griffith formula (14) in brittle fracture, one would have to establish a connection between the macroscopic strain and the microscopic strain at the tip of the crack. This would demand the solution of the plastic strain distribution problem in a body containing a crack of given length and orientation; strain hardening is probably an essential point which would have to be taken into account. The greatest complication, compared with the Griffith problem, would be the circumstance that the strain distribution around the crack in the ductile case will depend fundamentally on the shape of the body and on the distribution of the external forces acting upon it. The criterion of ductile fracture may turn out to involve no physical quantities apart from the yield stress–strain curve of the material; like the ultimate stress, the condition of ductile fracture may be derived in this case from the stress–strain distribution in the plastically strained body.

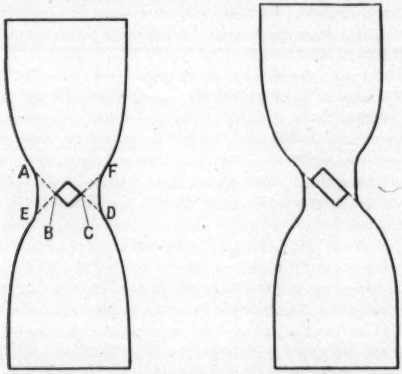
§34. TWO SPECIAL MECHANISMS OF DUCTILE FRACTURE

The general considerations of the preceding section can be illustrated by the consideration of two possible special mechanisms of ductile fracture.

Tensile fracture in flat specimens is frequently introduced by the formation of a tubular channel going right through the (elongated) neck; the formation of such a channel in the neck of an aluminium single crystal is shown by the photographs Figure 25 (*a*) and (*b*) (Plate III)*. A possible mechanism for the formation of the channel is indicated in Figure 26 (*a*). Let the dotted lines indicate slip planes belonging to two sets. If a small quadrangular channel is already present, plastic deformation needs the smallest tensile force if it takes place by glide along the parts AB and CD of the two slip planes running along two opposite sides of the channel. The operation of these two slip planes leads to a widening of the channel as shown in Figure 26 (*b*); if, at the same time, the slip planes EB and CF also operate, the cross section of the channel may remain similar to itself while it is enlarged by the slip process.

* The necked single crystal was kindly lent to the author by Dr C. F. Tipper–Elam; the photographs were taken by Mr. S. D. Charter.

The possible crack propagation process indicated in Figure 26(*a*), (*b*) shows strikingly that ductile fracture can take place without the cohesion being overcome by a cleavage-like process; if the channel formation occurs according to this



(*a*) (*b*)

Figure 26. Possible mechanism of crack propagation in neck.

mechanism, it is fundamentally a slipping-off process depending only on the plastic properties of the material. Naturally, the same type of process could occur with polycrystalline materials if the plastic deformation is concentrated in thin zones around planes of maximum shear stress, and it could also occur

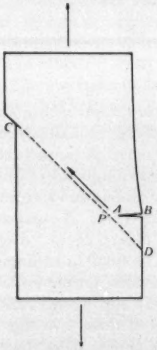


Figure 27. Possible ductile crack propagation mechanism.

in cylindrical specimens. It is very likely, therefore, that the mechanisms of ductile fracture arise directly from the plastic properties of the material and from the characteristics of the plastic deformation in the strained body.

A second possible mechanism of ductile fracture is shown in Figure 27. It was mentioned in §12 that the application of the Griffith formula to what seems to be the cleavage fracture of ductile crystals like NaCl or Zn would lead to the assumption of excessively long cracks. The mechanism indicated in Figure 27 (Orowan 1934b) shows how cleavage fracture might arise by plastic deformation even if the Griffith condition were not fulfilled. Let PAB be a cleavage plane and CD a slip plane in a crystal under the tensile stress indicated by the arrows, and A the tip of a surface crack in the cleavage plane. When a glide process starting at C arrives at the plane PAB, the stress in the small area PA must pull the triangular prism PABD along the slip plane PD in order to propagate the slip process to D; if PA is too small relatively to PD, cleavage along PA occurs instead of slip over PD, and the crack extends by the amount PA.

V. FATIGUE

§35. MECHANICAL FATIGUE UNDER CYCLIC STRESSING

According to Moore and Kommers (1927), the first investigation into fatigue of which there is a record is the work of Albert (1837) in 1829. Some of the principal laws of fatigue, including that of the existence of safe amplitudes, were discovered by Wöhler between 1860 and 1870.

The most important facts about mechanical fatigue can be summarized as follows:

(i) A metal may withstand a certain number of applications of a stress, or of a stress cycle, and break when this number is exceeded.

(ii) Fracture after cyclic stressing may occur even if the highest stress of the cycle is below the so-called elastic limit, i.e. the stress at which irreversible deformations become easily measurable.

(iii) Very often (with ferritic steels, almost always) there is a critical amplitude of alternating stress below which the material can withstand any number of cycles. Amplitudes below the critical are called safe, those above the critical, unsafe.

(iv) The critical amplitude is not influenced by the superposition of a constant stress (the "mean stress" of the cycle) if this is small compared with the elastic limit. In many cases, the critical amplitude for repeated loading (stress varying between zero and a maximum) is still approximately the same as for alternating load (stress varying between $+\sigma$ and $-\sigma$). With further increase of the mean stress, the critical amplitude decreases.

(v) If the logarithm of the stress amplitude S is plotted as a function of the logarithm of the number N of cycles after which fracture occurs, the resulting curve has usually the characteristic shape shown in Figure 28. The horizontal branch corresponds to the critical amplitude.

(vi) A material may show considerable damping (mechanical hysteresis and heat development) under safe cycles (Föppl 1936).

Some materials have no safe amplitudes; the usually horizontal branch of the $\log S$–$\log N$ curve is then slightly sloping downwards, while the general character of the curve remains the same. The absence of a safe amplitude may be due to metallurgical changes during the fatigue test; it has been suggested

(Hanstock 1948) that cyclic stressing of age-hardening aluminium alloys, which are noted for not having safe ranges in the strict sense of the word, causes a progressive precipitation.

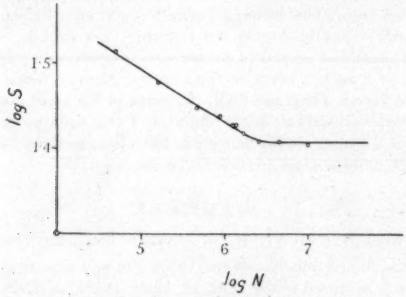


Figure 28. Log S–log N curve.

§36. THE NATURE OF FATIGUE FRACTURE

The first serious investigation into the mechanism of fatigue fracture was that by Ewing and Humfrey (1903); they observed microscopically that the crystal grains of Swedish iron developed slip bands under unsafe stress amplitudes, and that finally cracks opened up along the bands. Later, it was found that neither result can be generalized: slip bands appear very often under safe cycles, and the crack is frequently intercrystalline. This has been the fate of all other hypotheses on the fracture mechanism in fatigue: some appear plausible in certain cases, but are definitely not applicable to others. Eventually, it became clear that fatigue fracture, like fracture in general, is not one well-defined physical process; several essentially different mechanisms may lead to the formation and propagation of fatigue cracks. This is directly demonstrated by the fact, already mentioned, that the crack may be either transcrystalline or intergranular; transcrystalline cracks may possibly run either along slip planes or along cleavage planes, and the fracture may be accordingly either of the shear fracture or of the brittle fracture type. For this reason, the numerous hypotheses on the mechanism of fatigue fractures, many of them antiquated and most of them largely speculative, will not be discussed in this review.* The really important question is, how

* The reader may be referred to the books of Gough (1924), Moore and Kommers (1927), and Cazaud (1948). These monographs also contain detailed information about the technique of fatigue testing and the engineering aspects of fatigue. The most recent hypothesis, which has not received much discussion yet, is that of Dehlinger (1948 a, b) who assumed that the grains at the surface, not being supported by neighbouring grains from all sides, would suffer plastic deformation prematurely. Once this had taken place, the adjacent internal grains would become overloaded; under static load, they would undergo slow creep which would extend inwards by the overloading of the neighbouring grains. Dehlinger assumed that the same process (he uses the pictorial expression " squeezing out " of the surface grains) was the cause of fatigue. Fatigue cracks would be formed between the bent glide lamellae of the grains " squeezed out " from the surface; the critical stress amplitude would be the stress at which the first surface grains are " squeezed out ", and thus it would coincide with the absolute creep limit of the metal. Apart from the fact that this hypothesis does not really explain the phenomenon of fatigue and its main laws, its failure is obvious at once from the observation that single crystals show the same characteristic fatigue phenomena as polycrystalline metals (Gough and Hanson 1926).

physically different alternative mechanisms of fracture can result in the same fatigue behaviour, as summarized in the preceding section. The answer is (Orowan 1939) that the main features of fatigue are largely independent of details of the fracture mechanism because they arise from two general properties of metals: their structural inhomogeneities, and their capacity for strain-hardening. The importance of structural inhomogeneities was first emphasized by Gilchrist (1900) and then by Gough and Hanson (1923). The latter authors, in particular, recognized that the re-distribution of stresses in and around a plastic inclusion in an elastic bulk can lead to local increases of stress in the course of alternating stressing.

§37. THEORY OF MECHANICAL FATIGUE

Let it be assumed for simplicity that the plastic deformation of the metal shows no time effect (creep); the stress should rise with the total absolute amount of strain according to a stress–strain curve (which for alternating strain may not be the same as that for steadily increasing strain). This assumption is realistic since fatigue phenomena are hardly influenced by the frequency of the stress cycles.

If a material is ideally homogeneous, only elastic deformation can occur during the removal and re-application of a stress the first application of which may have caused plastic deformations. Since no irreversible structural changes can occur by purely elastic deformations, cyclic stressing of an ideally homogeneous material could not lead to fatigue. On the other hand, the application of plastic strain cycles of constant amplitude would result in the process represented in Figure 29

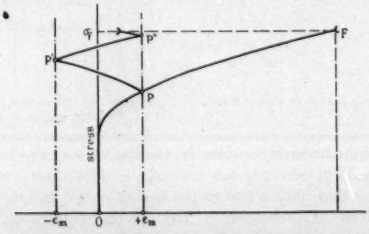


Figure 29. Diagram of fracture of homogeneous material under cycles of constant strain amplitude.

where ϵ_m is the amplitude of the plastic strain cycles. The stress would rise with increasing total strain, by strain hardening, according to a zig-zag curve whose pieces could be fitted together to a smooth curve OF representing the stress–strain curve for alternating straining with the given amplitude. With increasing number of cycles, the total plastic strain could increase to infinity, and the stress corresponding to it according to the stress–strain curve OF would also increase infinitely unless the curve became asymptotically horizontal. Consequently, fracture would always occur ultimately if its condition were that the total strain should attain a critical value; it would also occur if fracture obeyed a critical stress condition, unless the stress–strain curve converged to a horizontal below the value of the fracture stress.

An ideally homogeneous material, therefore, would show no fatigue under cycles of constant stress amplitude and, in general, no safe amplitude under cycles of constant plastic strain amplitude. If, therefore, real metals show both fatigue and safe amplitudes of stress and strain, this must be due to structural inhomogeneities. The salient point about inhomogeneities is that, if a mainly elastic body is subjected to cycles of constant stress amplitude, the plastic strain amplitude in small strain-hardening plastic inclusions in the body is not constant but diminishes in the course of the cycles. According to the fundamental theorem of fatigue (Orowan 1939), proved below, the total plastic strain in such an inclusion always converges towards a finite value unless the stress–strain curve becomes horizontal. This value increases with the amplitude of the stress applied to the body, and the amplitude is safe or unsafe according to whether the limiting total plastic strain in the inclusion is below or above the critical strain for fracture (if fracture obeys a stress criterion, the critical stress corresponds to a critical strain according to the stress–strain curve).

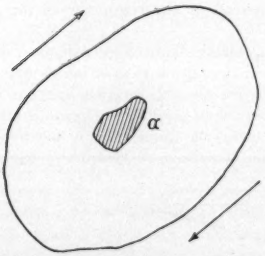


Figure 30. Plastic region in elastic matrix.

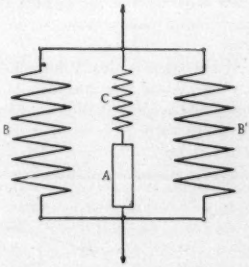


Figure 31. Model of the behaviour of a plastic region in elastic matrix.

To prove the fundamental theorem, let the shaded area α in Figure 30 represent a small plastic region in a mainly elastic body; and let a stress of such magnitude act upon the body that only a few of the softest spots, among them α, undergo plastic deformation. The volume α may be particularly soft because of the favourable orientation of its slip planes and directions, or because it is a region of stress concentration around a nearby crack or notch, or for any other reason. If the elastic stress in the bulk is given, the stress in the inclusion α depends on the plastic strain. The relationship between the stress and the plastic strain in α can be obtained simply by means of a model (Figure 31) representing the behaviour of a plastic inclusion in elastic surroundings. The two large springs B and B′ represent the elastic bulk to the sides of α; the small spring C in the centre stands for the elastic surroundings of α which connect it with the elastic bulk, and also for the elasticity of α itself which is superposed upon its plasticity. The purely plastic element A represents the plasticity of α, and the horizontal rigid bars replace those parts of the elastic bulk which are some distance above and below the level of α. The arrows indicate the applied stress.

An applied stress S causes an increase of length $\Delta = S/P$ of the large springs, where P is the spring constant for the two springs taken together; the small

spring C is too weak to influence Δ. If s is the stress in C and in the plastic element α, and p the spring constant of C, then the increase of length of C is s/p. Let now ϵ be the plastic elongation of A which corresponds to the stress s according to its stress–strain curve; consequently,

$$\Delta = S/P = s/p + \epsilon. \qquad \ldots\ldots (20)$$

For a constant applied stress S, this equation represents in the plastic stress–strain diagram for A a straight line BZ (Figure 32). $OB = S/P$ is the strain

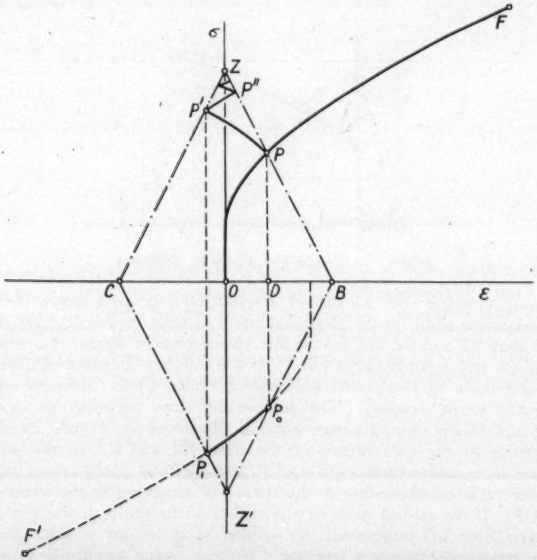


Figure 32. Diagram of strain hardening of a plastic region in an elastic matrix under a safe amplitude.

in the volume α if the stress in it vanishes, that is, if it is a hole or contains a liquid; $OZ = pS/P$ is the stress when the plastic strain in α vanishes. If the applied stress S is increased, the line BZ moves parallel to itself away from O.

While simultaneous values of s and ϵ in the plastic inclusion α must always lie on BZ, they must also lie on the stress–strain curve OF for the given amplitude of cyclic straining (Figure 32); the first application of the amplitude S_m of the external stress, therefore, will produce a stress and a plastic strain in α represented by the point of intersection P of BZ with the yield stress–strain curve OF. When the applied stress is removed, and an equal but opposite stress applied, the stress in α diminishes, while the plastic strain remains constant until the stress s has reached its previous magnitude with an opposite sign; i.e. as far as the point P_0. Then yielding begins with (algebraically) diminishing plastic strain, and

the stress–strain point moves along a stress–strain curve obtained by rotating PF through 180° around D. When the applied stress has reached the value $-S_m$, the point in the s–ϵ plot must lie both on the stress–strain curve P_0F' and on the line CZ′ that represents equation (20) for negative values of s and ϵ; thus the first reversal of stress will lead to the point P_1. Similarly, the second application of the positive amplitude gives plastic deformation beginning at P′ and ending at P″, and so on. The diagram can be simplified by folding its lower half upon the upper half around the ϵ axis; this results in Figure 33. In the course of

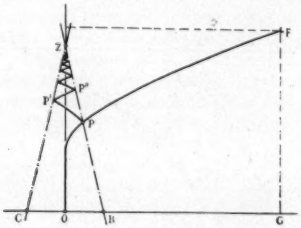


Figure 33. Diagram as in Figure 32, simplified.

alternating stressing with a constant amplitude of external stress, therefore, the stress–strain point for the plastic inclusion ascends between the converging straight lines BZ and CZ in a zig-zag line, the sections of which are consecutive parts of the stress–strain curve OF. As the number of cycles increases, the strain amplitude in the plastic inclusion decreases, while the stress and the total plastic strain increase. The decisive difference between this case and that of the ideally homogeneous material illustrated by Figure 29 is that here, owing to the convergence of the lines BZ and CZ in the point Z, the stress in α cannot exceed $OZ = pS/P$, and the total strain cannot increase above the value corresponding to the stress OZ according to the stress–strain curve OF. If the critical stress or the critical strain at which fracture occurs lie above OZ or OG respectively, no number of cycles can produce fracture, and the stress amplitude is a safe one. With an unsafe amplitude, illustrated by Figure 34, the critical stress or strain is reached below the point of convergence Z, and fracture occurs after a finite number of cycles.

Naturally, the fracture occurring at the point F′ in Figure 34 extends only over the small plastic region α; the resulting crack, however, creates a stress concentration and so a new plastic region next to α, in which the entire process is repeated. In this way the crack spreads until it reaches the length at which the next application of the full tensile stress of the cycle produces complete fracture. It is usually easy to recognize visually where the crack propagation has changed over from the fatigue to the static mechanism. Figure 35 (Plate III) shows a fragment "spalled" off a steel roll of a cold rolling mill; the fatigue fracture started at an inclusion visible at the centre of the concentric ellipse-like rings, and the change from fatigue to static fracture occurred at the border of the elliptical region. Up to the border, the crack propagation may have taken weeks; from the border onwards, a small fraction of a second.

For an unsafe amplitude, the relationship between the amplitude S of the applied stress and the number N of cycles to fracture can be recognized directly by counting the number of zig-zags of the curve in Figure 34 as a function of the height of the point Z which, as was seen, is proportional to the applied stress amplitude within the approximations made in this section. The number obtained

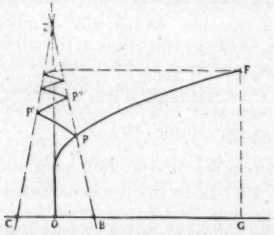


Figure 34. Diagram of fracture of a plastic region in an elastic matrix under cyclic stress of unsafe amplitude.

in this way is that for a single plastic region α; it may be proportional to the number needed for fracturing the specimen if the number of the consecutive small plastic regions across which the fatigue crack has to extend to produce complete fracture depends on the structure of the material rather than on the stress amplitude. The number of zig-zags to fracture depends somewhat on the shape of the stress–strain curve; however, this is not significant since, for not too high amplitudes,

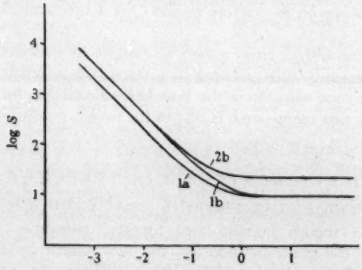


Figure 36. Log S–log N curves calculated from equation (21). For explanation of the curves see text.

most of the zig-zags lie in the upper part of the curve where this is approximately straight. By assuming a simple stress–strain curve with a sharp yield point and a straight part beyond the yield point, the relationship between S and N can be calculated; the result is (Orowan 1939)

$$N = C \log \frac{s_m - s_y}{s_m - s_f}, \qquad \ldots\ldots(21)$$

where C is a constant, $s_m = OZ = pS/P$ (Figures 33, 34), s_y the yield stress, and s_f the fracture stress. Curve 1*a* in Figure 36 represents equation (21) for the following values of the constants: $p/P = 1$, so that $S = s_m$; $s_y = 5$, $s_f = 10$, $C = 1$. Curve 1*b* is the same with $s_y = 0$; it shows how little the character of the curve depends even on the presence of a yield point. In curve 2*b*, thermal softening has been taken into account in a way described in the above-mentioned paper.

Figures 28 and 36 show that the calculated curves have the same form as the observed ones. For a complete, numerically quantitative theory, of course, knowledge of a number of factors (the shape of the stress–strain curve, the factor p/P, the critical stress or total strain for fracture etc.) would be necessary.

Figures 31 and 32 can easily be adapted for the case where the mean stress of the cycle is different from zero; clearly the critical amplitude remains the same, in accordance with the empirical rule (4), §35.

§38. CORROSION FATIGUE AND STRESS CORROSION

Haigh observed in 1917 (a, b) that cyclic stressing of certain brasses in a corrosive medium resulted in lowered values of the critical amplitude. This phenomenon was then investigated in detail by McAdam (1926, 1927), who found it to be very general; the critical amplitude is reduced only if cyclic stressing and the corrosive action are simultaneous. "Corrosion fatigue" usually begins with the formation of pits which later join together to rather coarse cracks. The metal is more anodic, and so more rapidly dissolved, at the bottom of the pits and cracks, because of the elastic stress concentration, or because of the latent energy of plastic distortion, or because plastic distortion currently ruptures protective films there (Evans 1948, Tchorabdji–Simnad and Evans 1947). This process slows down and finally stops when the cracks become too deep and sharp and the electric resistance of the corrosive solution in them becomes too high; they act, however, as "stress raisers", and mechanical fatigue takes over their propagation.

Stress corrosion differs from corrosion fatigue in that it can also take place under constant stress. The cracking of internally stressed brass by immersion in a mercurous nitrate solution is the best known example for this phenomenon which, in the case just mentioned, is due to the intercrystalline attack of mercury.

VI. INTERCRYSTALLINE FRACTURES

§39. INTERCRYSTALLINE VISCOUS FRACTURE (CREEP FRACTURE)

Polycrystalline metals contain two textural elements, the crystal grains (crystallites) and the grain boundaries, the mechanical behaviour of which is fundamentally different (Becker 1925, 1926). Plastic slip in the grains begins when the shear stress in a slip plane, resolved in a slip direction, reaches a more or less sharply defined critical value. Mechanical twinning, and kinking (Orowan 1942), are similarly bound to the attainment of a critical stress, although quantitative laws are not known in these cases. The critical shear stress for slip does not change rapidly with the temperature and remains, in general, finite up to the melting point (simplest example: melting ice is hard). It is not much influenced by the rate of deformation (example: the yield stress of copper in rapid compression when used as a "crusher" for measuring the gas pressure in guns is only some 10 or 20% higher than in slow compression).

On the other hand, the grain boundaries show all characteristic features of viscosity as produced by a disorderly arrangement of the atoms or molecules (Becker 1925, 1926, Orowan 1938, 1947). They are rigid below a comparatively narrow "softening range" of temperature; above it, they behave like a layer of a viscous liquid. They show no yield stress, and the rate of sliding of the grains upon their neighbours would probably be proportional to the shear stress if the boundaries were not uneven and the grains could slide without geometrical non-conformity. Owing to the "keying" of the grains to their neighbours, however, the rate of boundary sliding for a given low stress diminishes after a small deformation and may practically disappear. This effect was put to important practical use in the tungsten filaments of incandescent lamps before the modern long-grain filaments were available: by adding small amounts of certain oxides which precipitate in the grain boundary, secondary recrystallization was produced (Beck, Holzworth and Sperry 1948), and the typical ragged grain boundaries arising in this way "spliced" the grains together and reduced boundary creep.

If the stress is increased, it enforces sliding along the boundaries in spite of the geometrical non-conformity. As a consequence, cavities open up between the grains and finally fracture occurs (Hanson and Wheeler 1931). This is a typical occurrence in high-temperature creep tests where the specimen often breaks with apparently little deformation and no necking. The absence of necking is connected with the stability of extension of purely viscous rods (§ 5). The absence of considerable deformation in the grains is a consequence of the non-existence of a yield stress for boundary sliding; with increasing temperature, the typically amorphous boundaries soften much more rapidly than the grains which always retain a finite yield stress, and at high temperatures and low rates of straining, the grains may slide on one another as rigid units.

Since creep fracture is a consequence of viscous sliding, one cannot speak of a definite strength; the time needed for fracture increases with decreasing stress and temperature. If the boundary viscosity were Newtonian, the time needed for fracture would be inversely proportional to the stress; actually, it decreases more rapidly with increasing stress and, in many cases, an approximately parabolic relationship is observed in limited ranges of temperature. If the logarithm of the time of fracture is plotted as a function of the logarithm of stress, the curve is often found to consist of two straight lines (Siegfried 1943, Parker 1944); the line corresponding to lower stresses is interpreted as representing fracture by the viscous sliding of the grains as solid blocks, while in the stress range of the other line transcrystalline slip and fracture occur.

§40. INTERCRYSTALLINE BRITTLE FRACTURE

There are numerous instances of chemical processes in the grain boundaries so reducing the intercrystalline cohesion that the boundaries become the weakest element in the metal; when fracture occurs, it runs along them. Several cases have already been mentioned, such as the temper-brittleness of steel, and the stress corrosion in brasses. A well known further example is the "burning" of steel heated at excessive temperatures; formerly it was thought that burning is a grain boundary oxidation, but Jominy (1929) has been able to produce it *in vacuo*, and so its chemism must be more complex.

VII. FRACTURE BY MOLECULAR SLIDING AND SPECIAL TYPES OF FRACTURES

§41. FRACTURE BY MOLECULAR SLIDING

This mechanism occurs with organic high polymers; it is very similar in principle to creep fracture (§39), but the structural units which slide on each other are here not crystal grains but long molecules. Like creep fracture, it is not a pure crack propagation process: the formation of the fracture crack is preceded by a diminution of the cohesion of the whole stressed body by the formation of cavities (which are here mostly submicroscopic, instead of microscopic), and the fragments are permanently weaker than the initial material. The mechanism of the process is so intimately connected with the chemical characteristics of the material that a general treatment is hardly possible; for details, the special literature should be consulted (Alfrey 1948, Busse *et al.* 1942, Eley and Pepper 1947).

§42. SPECIAL TYPES OF FRACTURE

The classification used in this review is not intended to imply that most fracture phenomena have a natural place in one of its pigeon-holes. There are many cases where the mechanism of fracture is closely linked up with some very special textural or structural feature of a material. As an example, the fracture of dendritic materials and cast irons may be mentioned (Iitaka 1931, Iitaka and Yamagishi 1938). In such cases it is of particular importance to concentrate attention upon the microscopic or submicroscopic features of the fracture. In recent years, both the microscopic and electron-microscopic techniques of the study of fractures have made great progress (Zapffe 1948, Zapffe, Landgraf and Worden 1948, Nutting 1949), and it is more imperative than ever not to neglect direct observation by relying too much on deductions from mechanical tests.

ACKNOWLEDGMENT

The author is indebted to Dr. J. F. Nye for reading several parts of the manuscript of this review.

REFERENCES

ALBERT, —, 1837, *Arch. Mineral., Geognosie, Bergbau u. Hüttenk.*, p. 215; see also *Stahl u. Eisen*, 1886, **438,** 496.
ALFREY, T., Jr., 1948, *Mechanical Behaviour of High Polymers* (New York).
ANDRADE, E. N. DA C., and MARTINDALE, J. G., 1935, *Phil. Trans. Roy. Soc.* A, **235,** 69.
ANDRADE, E. N. DA C., and TSIEN, L. C., 1937, *Proc. Roy. Soc.* A, **159,** 346.
BAKER, T. C., and PRESTON, F. W., 1946, *J. Appl. Phys.*, **17,** 179.
BECK, P. A., HOLZWORTH, M. L., and SPERRY, P. R., 1948, *Amer. Inst. Mech. Engrs.*, Pub. No. 2475, *Met. Tech.*, September.
BECKER, R., 1925, *Phys. Z.*, **26,** 919; 1926, *Z. tech. Phys.*, **7,** 547.
BRAGG, W. L., 1942, *Nature, Lond.*, **149,** 511.
BRAGG, W. L., and LOMER, W. M., 1949, *Proc. Roy. Soc.* A (in press).
BRIDGMAN, P. W., 1944, *Trans. Amer. Soc. Met.*, **32,** 533; 1947a, *Rev. Mod. Phys.*, **17,** 3; 1947b, *J. Appl. Phys.*, **18,** 246.
BUSSE, W. F., LESSIG, E. T., LOUGHBOROUGH, D. L., and LARRICK, L., 1942, *J. Appl. Phys.*, **13,** 715.
CAZAUD, R., 1948, *La fatigue des métaux*, 3rd ed., Paris.
CONSIDÈRE, A., 1885, *Ann. des ponts et chaussées*, **6,** 9.
DAVIDENKOV, N. N., 1936, *Dinamicheskaya ispytania metallov*, Moscow.
DEHLINGER, U., 1940a, *Z. Phys.*, **115,** 625; 1940b, *Z. Metallkde.*, **32,** 199.
ELEY, D. D., and PEPPER, D. C., 1947, *Trans. Faraday Soc.*, **43,** 559.

ELLIOTT, H. A., 1947, *Proc. Phys. Soc.*, **59**, 208.
EPSTEIN, B., 1947, *J. Franklin Inst.*, **244**, 471.
EVANS, U. R., 1948, *An Introduction to Metallic Corrosion* (London: Edward Arnold).
EWALD, W., and POLANYI, M., 1925, *Z. Phys.*, **31**, 746.
EWING, Sir J. A., and HUMFREY, J. C. W., 1903, *Phil. Trans. Roy. Soc.* A, **200**, 241.
FETTWEIS, F., 1929, *Arch. Eisenhüttenwesen*, **2**, 625.
FISHER, J. C., and HOLLOMON, J. H., 1947, *Amer. Inst. Mech. Engrs.*, Tech. Publ. No. 2218; *Met. Tech.*, **14**, 5.
FLOREŃSKI, —, 1930, *Vestnik Elektrotekh.*, No. 2 (Russian).
FÖPPL, O., 1936, *J. Iron & Steel Inst.*, **134**, 393.
FRENKEL, J., 1926, *Z. Phys.*, **37**, 572.
FRENKEL, J., and KONTOROVA, T. A., 1943, *J. Phys. U.S.S.R.*, **1**, 108.
FRIDMAN, J. B., 1941, *Zhurn. Techn. Fiziki*, **11**, 985 (Russian); 1946, *Deformatsia i razrushenie metallov pri staticheskikh i udarnykh nagruzkakh*, Moscow (Russian).
GENSAMER, M., 1941, *Strength of Metals under Combined Stresses* (Cleveland), p. 55.
GILCHRIST, J., 1900, *Engineer* (*London*), **90**, 203.
GOUGH, H. J., 1924, *The Fatigue of Metals*, London.
GOUGH, H. J., and HANSON, D., 1923, *Proc. Roy. Soc.* A, **104**, 538; 1926, *Phil. Trans. Roy. Soc.* A, **226**, 1.
GRIFFITH, A. A., 1920, *Phil. Trans. Roy. Soc.* A, **221**, 163; 1924, *First Internat. Congr. Appl. Mech* (Delft), 55.
HAIGH, B. P., 1917a, *Journ. Inst. Met.*, **18**, 55; 1917b, *Engineering* (London), **21**, 315.
HANSON, D., and WHEELER, M. A., 1931, *J. Inst. Met.*, **45**, 229.
HANSTOCK, R. F., 1948, *J. Inst. Met.*, **74**, 469.
HENCKY, H., 1923, *Z. angew. math. Mech.*, **3**, 241.
HILL, R., 1949, *Quart. J. Mech. Appl. Math.*, **5** (in press).
HOLLAND, A. J., and TURNER, W. E. S., 1940, *Trans. Soc. Glass Techn.*, **24**, 46.
HOLLOMON, J. H., 1946, *The Problem of Fracture* (New York).
HOLLOMON, J. H., and ZENER, C., 1944, *Trans. Amer. Inst. Mech. Engrs.*, **158**, 283.
HOYT, S. L., 1936, *Metals and Alloys*, **7**, 5, 39, 102, 140.
IITAKA, I., 1931, *Proc. Imp. Acad.* (Japan), **7**, 337.
IITAKA, I., and YAMAGISHI, I., 1938, *Sci. Pap. Inst. Phys. Chem. Res.*, **34**, 1025.
INGLIS, C. E., 1913, *Trans. Inst. Naval Archit.*, **55**, pt. I, 219.
JENCKEL, E., 1932, *Z. Elektrochem.*, **38**, 569.
JOFFÉ, A. F., 1928, *The Physics of Crystals* (New York).
JOFFÉ, A., KIRPITSCHEWA, M. W., and LEVITZKY, M. A., 1924, *Z. Phys.*, **22**, 286.
JOMINY, W. E., 1929, *Trans. Amer. Soc. Steel Treating*, **16**, 298.
VON KÁRMÁN, TH., 1911, *Z. Ver. deutsch. Ing.*, **55**, 1749.
KARMARSCH, K., 1858, *Mitt. gew. Ver. Hanover*, 138.
KICK, F., 1892, *Z. Ver. deutsch. Ing.*, pp. 278 and 919.
KONTOROVA, T. A., 1940, *J. Tech. Phys. U.S.S.R.*, **10**, 886.
KUNTZE, W., 1928, *Arch. Eisenhüttenwesen*, **2**, 109; 1931, *Z. Phys.*, **72**, 785; 1932, *Kohäsionsfestigkeit* (Berlin)
LUDWIK, P., 1926, *Z. Metallkde.*, **18**, 269.
LUDWIK, P., and SCHEU, R., 1923, *Stahl u. Eisen*, **43**, 999.
MCADAM, D. J. Jr., 1926, *Proc. Amer. Soc. Testing Materials*, **26**, II, 224; 1927, *Trans. Amer. Soc. Steel Treating*; 1941, *J. Appl. Mech.* **8**, A 155; 1942, *Amer. Inst. Mech. Engrs.*, Tech. Publ. 1414 (Met. Tech.).
MCADAM, D. J., Jr., and MEBS, R. W., 1943, *Amer. Inst. Mech. Engrs.*, Tech. Publ. 1615 (Met. Tech., August).
MAIER, A. F., 1935, *Einfluss des Spannungszustandes auf das Formänderungsvermögen der metallischen Werkstoffe*, *V.D.I.*, Berlin.
MOORE, H. F., and KOMMERS, J. B., 1927, *The Fatigue of Metals*, New York.
NADAI, A., 1931, *Plasticity*, New York.
NADAI, A., and MANJOINE, M. J., 1941, *J. Appl. Mech.*, A-77.
NUTTING, J., 1949, (Cavendish Laboratory, Cambridge, and British Iron and Steel Res. Assoc.); to be published.
OBREIMOFF, J. W., 1930, *Proc. Roy. Soc.* A, **127**, 290.

OROWAN, E., 1933a, *Z. Phys.*, **82,** 235; 1933b, *Ibid.*, **86,** 195; 1934a, *Z. Kristallogr.* (A), **89,** 327; 1934b, *Internat. Conf. on Phys.*, London, Vol. II, 81; 1938, *Proc. Roy. Soc.* A, **168,** 307; 1939, *Ibid.*, **171,** 79; 1940, *Proc. Phys. Soc.*, **52,** 8; 1942, *Nature, Lond.*, **149,** 643; 1944, *Ibid.*, **154,** 341; 1945, *Trans. Inst. Eng. and Shipbuild.*, Scotland, 165; 1947, *West of Scotland Iron and Steel Inst.*, 45; 1949a, *The Principles of Rheological Measurement*, (Edinburgh), article on *Mechanical Testing of Solids* (*Conf. Brit. Rheol. Club*, Oct. 1946; 1949b, to be published.

OROWAN, E., NYE, J. F., and CAIRNS, W. J., 1945, *Theoret. Res. Rep.* No. 16/45, *Arm. Res. Dept., M.O.S.*, London.

PARKER, E. R., 1944, *Trans. Amer. Soc. Metals*, **33,** 150.

PEIRCE, F. T., 1926, *J. Text. Inst.*, **17,** 355.

PETCH, N. J., 1949 (Cavendish Laboratory, Cambridge, and British Iron and Steel Res. Assoc.), to be published.

POLANYI, M., 1921, *Z. Phys.*, **7,** 323.

PRANDTL, L., 1923, *Z. angew. Math. Mech.*, **3,** 401.

PRESTON, F. W., 1942, *J. Appl. Phys.* **13,** 623.

REINKOBER, O., 1931, *Phys. Z.*, **32,** 243; 1932, *Ibid.*, **33,** 32.

RUSSELL, S. B., 1897, *Proc. Amer. Soc. Civ. Engrs.*, **23,** 550.

SACHS, G., and LUBAHN, J. D., 1942, *Iron Age*, **150,** 31, 49.

SACK, R. A., 1946, *Proc. Phys. Soc.*, **58,** 729.

SAKHAROV, P. S., 1936, *J. Tech. Phys. U.S.S.R.*, **6,** 1381.

SCHMID, E., and VAUPEL, O., 1929, *Z. Phys.*, **56,** 308.

SCHURKOW, S., 1932, *Phys. Z. Sowjet*, **1,** 132.

SHEARIN, R. E., RUARK, A. E., and TRIMBLE, R. M., 1948, *Trans. Amer. Soc. Metals.*

SIEBEL, E., 1925, *Berichte der Fachausschüsse des Vereins deutscher Eisenhüttenleute, Werkstoffausschuss*, Ber. No. 71.

SIEGFRIED, W., 1943, *Trans. Amer. Inst. Mech. Engrs.*, **65,** A 202.

SMEKAL, A., 1927, *Ann. Phys. Lpz.*, **83,** 1202.

SOHNCKE, L. 1869, (Poggend.), *Ann. Phys. Lpz.*, **137,** 177.

STANLEY-SMITH, G., 1945, *Metallurgia*, **32,** 163.

SWIFT, H. W., 1939, *J. Iron & Steel Inst.*, **140,** 181.

TCHORABDJI-SIMNAD, M., and EVANS, U. R., 1947, *Proc. Roy. Soc.* A, **188,** 372.

WEIBULL, W., 1939a, *Ingen. Vetensk. Akad.* (Stockholm), Proc. No. 151; 1939b, *Ibid.*, Proc. No. 153.

WÖHLER, A., 1860–1870, *Z. Bauwesen*, vols. between 1860 and 1870.

YOUNG, T., 1805, *On the Cohesion of Fluids*, *Phil. Mag.* See also Lord RAYLEIGH, *Collected Works*, Vol. III, p. 397.

ZAPFFE, C. A., 1948, *Iron Age*, April 1.

ZAPFFE, C. A., LANDGRAF, F. K., Jr., and WORDEN, C. O., Jr., 1948, *Science*, **107,** 320.

ZSCHOKKE, H., and MONTANDON, R., 1944, *Schweiz. Arch. angew. Wiss. u. Tech.*, **10,** 129.

ZWICKY, F., 1923, *Phys. Z.*, **24,** 131.

MULTIPOLE RADIATION IN ATOMIC SPECTRA

By A. RUBINOWICZ
Institute of Theoretical Mechanics, University of Warsaw

CONTENTS

PAGE
§ 1. Nomenclature and definitions 233
§ 2. Identification of new multipole lines 234
§ 3. Methods for the experimental determination of the multipole character of sources of light 235
§ 4. Zeeman effect of mixed electric quadrupole and magnetic dipole radiations 237
§ 5. Hyperfine structure of the electric quadrupole and magnetic dipole spectral lines 239
§ 6. The transition probabilities 240
§ 7. The electric octupole and magnetic quadrupole radiations 241

ABSTRACT. It is the purpose of this Report to summarize the more important of the recent advances in the field of atomic electromagnetic multipole radiation. The main stress is laid on results for the optical spectrum; but due regard is also paid to the most important developments in other spectral regions. Less attention, however, is given to subjects treated in the older reports on multipole radiation (Segrè 1931, Rubinowicz and Blaton 1932, Bowen 1936, Niewodniczański 1936, Mrozowski 1944). The multipole radiation of molecules is not discussed at all.

§ 1. NOMENCLATURE AND DEFINITIONS

All spectral lines which do not obey the selection rules for spontaneous electric dipole radiation are now called "forbidden" lines. Among them we have to distinguish two groups of different origin:

(i) Lines enforced by external influences such as external electromagnetic fields and neighbouring atoms. These lines are due to enforced electric dipole transitions.

(ii) Spontaneous (non-enforced) spectral lines due to higher multipole transitions, discovered for the first time in 1928 by I. S. Bowen (1936) in the planetary nebulae.

But some of the selection rules for spontaneous electric dipole radiation depend upon various factors, such as the coupling between the electrons constituting the atom (*LS*-, *jj*- or intermediate coupling cases) and the nuclear spin, so that the name "forbidden lines" does not correspond to any exactly defined concept. A proposition to base the concept of "forbidden lines" on the smallness of the corresponding transitions (Mrozowski 1944) does not seem more acceptable. Therefore we shall not follow further Bowen's usage and shall not refer to the higher multipole lines as forbidden lines.

The different types of multipole radiation are, however, precisely defined by the form of the variation with direction in space of the intensity and polarization of an electromagnetic radiation (Hansen 1935, Heitler 1936, Blaton 1937, Kramers 1938, Dancoff and Morrison 1939; cf. also more recent papers on multipole radiation by Yvon 1943, Humblet 1944a, b, 1946, Durand 1946a, b, Pluvinage 1947, Araki 1948).

A deeper insight into the mathematical structure of the electric and magnetic multipole radiation fields is given according to Kramers (1943) by group theory. From an electromagnetic radiation field $\mathscr{M}$ we can obtain generally a new solution $\mathscr{M}'$ of Maxwell's equations by applying to the original solution $\mathscr{M}$ an arbitrary rotation about some fixed point in space. We can then seek a finite set of linearly independent solutions $\mathscr{M}_1, \mathscr{M}_2, \ldots \mathscr{M}_g$ such that the rotated solutions $\mathscr{M}'_1, \mathscr{M}'_2, \ldots \mathscr{M}'_g$ can be expressed linearly in terms of the original $\mathscr{M}$'s in a way which is irreducible with respect to the complete group of rotations. The electromagnetic multipole radiation fields give the answer. For each positive value of the integer l there exist two sets of $2l+1$ linearly independent fields $\mathscr{M}_i^{(e)}$ and $\mathscr{M}_i^{(m)}$, such that every rotated $\mathscr{M}'^{(e)}_i$ and $\mathscr{M}'^{(m)}_i$ can be expressed by the corresponding original set $\mathscr{M}_i^{(e)}$ or $\mathscr{M}_i^{(m)}$ respectively. If we choose these fields in such a way that they behave differently with respect to reflection relative to the centre we get the electric and the magnetic 2^l-pole radiation fields respectively.

The name multipole comes from potential theory, where it is associated with an intuitive geometrical picture. We obtain a multipole source of order 2^l if we take a multipole source of order 2^{l-1} and place opposite to it a multipole source of the same order obtained by a translation of the first source and a change in the sign of its strength; that is, we get a multipole source of order 2^l by a directional derivation of the source coordinates of a multipole source of the next lower order 2^{l-1}.

In the case of the wave equation $\Delta u + k^2 u = 0$ this process of derivation of source coordinates will usually turn a multipole source of order 2^{l-1} into a system of two sources of order 2^l and 2^{l-2} respectively. In the domain of the wave equation it is therefore in general impossible to imagine that a multipole of the order 2^l consists of 2^l simple sources. The process of superposition of 2^l simple sources described above would lead to a general source system containing multipole sources of the orders 2^l, 2^{l-2}, 2^{l-4}, . . .

In the case of Maxwell's equations, the situation is even more complicated. A directional derivation of an electric multipole source of order 2^l, for instance, gives an electromagnetic field generated usually by two such multipole sources of orders 2^{l+1} and 2^{l-1} and a magnetic multipole source of order 2^l. The name multipole source seems, therefore, in the case of these differential equations, to be in some respects misleading. Using the ordinary definition of multipole sources we cannot say in general that a multipole source of the order 2^l consists of 2^l simple sources arranged in a certain manner.

In the following considerations we shall refer to the electric quadrupole and magnetic dipole radiations taken together as the electromagnetic multipole radiation of the second class, and to the electric octupole and magnetic quadrupole radiations as the electromagnetic multipole radiation of the third class etc.

§ 2. IDENTIFICATION OF NEW MULTIPOLE LINES

The most important achievement we have to record in the field of identification of new multipole lines is the solution of the old riddle of the origin of the coronal lines. So far not one of these spectral lines, of which about twenty are known, has been obtained in laboratory sources of light. Their identification was made possible by the investigations of the spectra of highly ionized atoms which we owe to Bengt Edlén. The starting point was Grotrian's discovery (in 1937 : see Edlén 1945) that the wave numbers of the strong coronal lines λ 6374·71

and λ 7891·6 coincide with some of the level separations within the ground terms of the spectra of Fe X and Fe XI. Following the same line of thought, Edlén (1942, 1945 : see also Swings 1945, Bowen and Swings 1947) thereupon associated two further coronal lines with transitions between ground terms in the Ca XII and Ca XIII spectra. The identification of the remaining coronal lines had to be based on extrapolations of the wave numbers corresponding to a given transition between two fixed levels (e.g. $^2P_{\frac{3}{2}} \rightleftharpoons {}^2P_{\frac{1}{2}}$) within the ground terms (e.g. $3s^23p$) in sequences of isoelectronic ions (e.g. Al I, Si II, P III, . . .). In this way it was possible to identify 19 spectral lines out of a list of 23 coronal lines as given by Grotrian and Lyot. The identifications were confirmed by a calculation of the relative intensities of the identified spectral lines under certain simplified conditions. The strongest coronal lines are iron lines.

We do not intend to give here a complete account of all the other multipole lines so far observed, because these are principally of astrophysical interest. But we have to mention a new field of spectroscopy in the initial stage of its development in which multipole radiations of the second class, especially magnetic dipole radiation, play an important rôle. This is radio-frequency and microwave spectroscopy applied to transitions between two Zeeman levels of the nucleus or between two Zeeman levels of the same fine or hyperfine structure term. Although this new branch of spectroscopy uses non-optical observation methods, on account of certain specific advantageous experimental circumstances (large intensities and exactly defined wavelengths of the incident radiation and a high sensitivity in the methods of measurement compensate for small transition probabilities), it is certainly destined to make large contributions to the development of optical spectroscopy. The method of radio-frequency spectroscopy was developed in 1938 by Rabi, Zacharias, Millman and Kusch (see Kellogg and Millman 1946), and the first application of microwave spectroscopic methods to an atomic problem was made in 1946 by Roberts, Beers and Hill (1946: see also Bleaney 1946, Gordy 1948).

§ 3. METHODS FOR THE EXPERIMENTAL DETERMINATION OF THE MULTIPOLE CHARACTER OF SOURCES OF LIGHT

The multipole character of an optical source of light can be determined experimentally by means of the Zeeman effect. Zeeman's early experiments proved that the radiation ordinarily emitted by atoms is an electric dipole radiation. It is also through this effect that the electric quadrupole and magnetic dipole radiations were discovered (Frerichs and Campbell 1930, Segrè and Bakker 1931, Niewodniczański 1933, 1934 a, b, Blaton, Niewodniczański 1934). In special cases the Stark effect might also be used for the same purpose.

To fix the multipole character of a line we can in some cases take advantage of the fact that a crystal lattice immobilizes to some extent the orientations of the atoms of which it is built up. But generally only those atoms in crystal lattices will give sharp spectral lines which contain a deeper lying uncompleted inner subshell, and that only in the case of transitions between levels which belong to the electron configuration of this subshell. Since all such levels are of the same parity, only multipole radiations of the second class can appear here in spontaneous transition according to the generalized Laporte rule. Since all these levels belong to the same electron configuration, the magnetic dipole radiation, even

though it can appear in the optical spectrum only in intercombination lines, will ordinarily be stronger than the electric quadrupole radiation.

Of all the spectra belonging to this group only the multipole character of the lines of the rare earths has been determined experimentally. Deutschbein (1940) examined the radiation of Eu^{+++} ions in europium ethyl sulphate. This compound crystallizes in the hexagonal system. Observing at right angles to the direction of the principal axis (c-axis) we get a spectrum consisting of lines polarized in a plane parallel or perpendicular to the direction of the principal axis. Assuming that the radiation is due to either electric or magnetic dipoles, the appearance or disappearance of a spectral line when observed from a direction parallel to the principal axis indicates whether it is due to an electric or magnetic dipole. In this way Deutschbein was able in the case of Eu^{+++} ions to establish the existence of the magnetic and the (evidently enforced) electric dipole radiations. Rosa (1943) found both types of spectral lines also in the radiation of Dy^{+++} ions.

The question of the conclusions to be drawn from such measurements using crystals belonging to any crystal system was discussed by Hellwege (1942). Like Deutschbein, he assumes that the radiation comes either from electric or magnetic dipoles which have fixed orientations relative to the lattice, and that we observe the radiation in the three directions given by the axes of the indicatrix of the crystal. Only in cases where the dipoles have favourable orientations with respect to the principal axes of the indicatrix can we determine their orientations in space, otherwise only "revolving" dipoles seem to appear. This supports the conclusion that only the complete examination of the directional dependence of the intensity and the polarization distributions of the radiation can uniquely determine its multipole character. In a further paper, Hellwege (1943) has extended these considerations, and in particular has shown how to calculate transition probabilities of the electric and magnetic dipoles from such measurements.

The radiation of the other rare earth ions examined (e.g. Pr^{+++} and Nd^{+++}, Hellwege (1941)) corresponds to fixed or revolving electric dipoles. Special attention may be drawn to the fact that in the experimental examination of the radiations of the rare earth ions no indication of the appearance of the electric quadrupole radiation has so far been found in any case.

These experimental results agree very well with the theoretical investigations carried out by Broer, Gorter and Hoogschagen (1945) and Broer (1946) in continuation of the earlier work of Bethe (1929), Bethe and Spedding (1937), Spedding (1940) and especially of Van Vleck (1937). Broer and his collaborators have very carefully estimated the oscillator strengths for the second-class multipole radiation of the rare earth ions. They find oscillator strengths for the electric quadrupole radiation of the order of 10^{-9}, which is decisively outside the order of 10^{-7} to 10^{-5} for the observed oscillator strengths of the rare earth ions. On the other hand the appearance of the magnetic dipole radiation is consistent with the estimates of these authors. Besides the magnetic dipole lines observed so far, such lines should also appear in the spectra of Gd^{+++}, Tb^{+++} and Ho^{+++}.

We can also determine the multiple character of the radiation of a source of light using the method of wide-angle interference. Its main feature is that it does not require fixation of the orientation of the radiating atom, thus being applicable, for example, to ions in solution. This method uses two beams issuing from a source of light examined at a wide angle 2ϕ, a direct beam and a beam reflected from a mirror. It had already been used by Selényi (1911), who proved

with its help that the atoms in a fluorescent solution emit an electric dipole radiation. Halpern and Doermann (1937, 1939 : see also Doermann 1938, Doermann and Halpern 1939, Selényi 1939) have then theoretically examined how the visibility of interference fringes for each kind of multipole radiation varies with the angle of divergence, 2ϕ, of the beams, with their polarization and with the experimental arrangement in general. Freed and Weissman (1941) and Freed (1942) have used this method in the case of Eu^{+++} ions in different solutions to ascertain the multipole character of their spectral lines. They found in the fluorescent spectrum of Eu^{+++} three groups of lines which according to Gobrecht (1937) correspond to the $J \to J'$ transitions $0 \to 0, 0 \to 1$ and $0 \to 2$. The experiments showed that the transition $0 \to 1$ is practically a magnetic dipole transition and the transitions $0 \to 0$ and $0 \to 2$ are (enforced) electric dipole transitions.

The method of wide-angle interference for the determination of the multipole character of spectral lines seems to be of great importance for a spectral region where the Zeeman effect is not applicable, viz. for x-rays. Mirrors of course do not exist here. But we can use the interference of a direct beam issuing from an atom in the lattice with beams scattered by its neighbouring atoms. In this way we obtain the complete reflection system of a crystal lattice. This was discovered by Kossel, but so far used only in the case of electric dipole radiation. For atoms of high atomic number the intensity of the electric quadrupole radiation attains nearly one-tenth of the maximum intensity of the electric dipole radiation in the x-ray region ; it may therefore be possible in this region to establish experimentally the existence of the electric quadrupole radiation. The theory of this effect can be given without difficulty by a generalization of Laue's (1941) theory for the electric dipole radiation. Complete success can of course only be expected after removing all the experimental and theoretical difficulties which are still present in the case of the electric dipole radiation.

§4. ZEEMAN EFFECT OF MIXED ELECTRIC QUADRUPOLE AND MAGNETIC DIPOLE RADIATIONS

Let us assume in classical electrodynamics a given distribution of electric currents emitting monochromatic radiation. If we split this up into the different multipole radiations, they will have constant phase relations and will therefore interfere. Two light waves having for given directions of propagation and polarization the amplitudes A and a and a phase difference ϕ give a resultant intensity proportional to $A^2 + a^2 + 2Aa\cos\phi$, and not to $A^2 + a^2$ as in the case of non-coherent waves. Nevertheless, in the expression for the total emitted energy there are no interference terms. Their contributions to the intensities in different directions reduce on the whole to zero because the phases ϕ are dependent on the direction of propagation.

As may be expected from the correspondence principle, we meet the same state of affairs in the quantum theory of radiation. First, Milianczuk (1935) and then Schiff (1940) (cf. Jenkins and Mrozowski 1941 b) called attention to the fact that the existence of the interference effect of the simultaneous electric quadrupole and magnetic dipole radiations can be established with the aid of the Zeeman effect where the directional dependence of the intensity of the radiation can be examined. The quantum theory of this effect was given by Milianczuk (1935) and in a more general form by Gerjuoy (1941) and by Shortley and his associates (1941). Its existence was proved by Jenkins and Mrozowski (1941 b),

who used for this purpose the Pb I spectral line $\lambda 7330{\cdot}12$. Both the initial state $^1D_2$ of this line and its final state $^3P_1$ belong to the same electron configuration $6s^26p^2$. Therefore the contribution of the magnetic dipole radiation to the intensity of this line is greater than that of the electric quadrupole radiation. In fact the admixture of this last radiation was determined to be only 2%, so that on an average $a/A = 0{\cdot}14$ if A and a are the amplitudes of the magnetic dipole and electric quadrupole radiations respectively. Despite the smallness of this admixture we have here easily observable effects because the departures from the intensity distribution of the magnetic dipole radiation A^2 are caused by the interference term $2Aa\cos\phi$ and not by the electric quadrupole term a^2. The ratios of the interference term and of the electric quadrupole radiation term to the intensity term of the magnetic dipole radiation are of the order $a/A = 0{\cdot}14$ and $a^2/A^2 = 0{\cdot}02$ respectively. The influence of the smaller amplitude a is increased in the interference term by the presence of the greater amplitude A. Such interference terms appear of course only in those Zeeman components which are allowed by the selection rules for both the multipole radiations of the second class.

How much the spectral line $\lambda 7330{\cdot}12$ contains of the magnetic dipole and electric quadrupole radiations can also be determined, according to Gerjuoy, from the intensity ratio of the two spectral lines $\lambda 4618$ and $\lambda 5313$ having the common initial level $^1S_0$. It is surprising that one finds in this way 6% for the admixture of the electric quadrupole radiation. The contradiction between the two determinations is not yet explained. Mrozowski (1944) considers it necessary to perform new precise intensity measurements of the interference effect in the line $\lambda 7330{\cdot}12$. for transverse and for longitudinal observations.

The interference effect may also be used in some cases for the measurement of the intensity of a multipole radiation of the second class which is masked by the background in the spectrum of the spectrograph. If we have, for instance, a spontaneous electric quadrupole radiation of amplitude a which is considerably smaller than the amplitude b corresponding to the background radiation, then we can apply an external electric field and measure the Zeeman effect of the mixed enforced electric dipole and the spontaneous electric quadrupole radiations. If A is the amplitude of the enforced electric dipole radiation and ϕ the phase difference between this radiation and the spontaneous electric quadrupole radiation, the resulting energy is proportional to $A^2 + 2Aa\cos\phi + a^2 + b^2$. Let us suppose that the intensity of the spontaneous electric quadrupole radiation is n times smaller than that of the background, i.e. that $a^2 = b^2/n$. In order to be measured, the contribution of the interference term must be stronger than that of the background radiation term, so that $2Aa\cos\phi > b^2 = na^2$ or $a/A < (2\cos\phi)/n$. This means that the amplitude A of the enforced dipole radiation should not be too small in comparison with the amplitude a of the spontaneous electric quadrupole radiation. But for an exact determination of the interference term its ratio to the enforced dipole term $(2Aa\cos\phi)/A^2$ should be as high as possible. With regard to the above inequality we have however

$$\frac{2Aa\cos\phi}{A^2} < \frac{4\cos^2\phi}{n},$$

so that the accuracy of the determination of the interference term is, under given circumstances, the lower the greater is the ratio n of the intensity of the background radiation to that of the spontaneous electric quadrupole radiation.

§5. HYPERFINE STRUCTURE OF THE ELECTRIC QUADRUPOLE AND MAGNETIC DIPOLE SPECTRAL LINES

By a careful choice of experimental conditions it was possible to increase the intensity of the multipole lines of the second class so that not only their hyperfine structure (Mrowozski 1940 a, b, 1946) but also the Zeeman effect in the hyperfine structure (Jenkins and Mrozowski 1941 a, b) could be investigated.

In the case of the electric dipole lines we get the selection rules and the intensity formulae for the hyperfine structure components from the corresponding rules and formulae for the fine structure multiplet and its Zeeman effect components simply by replacing the quantum numbers L, S and J by J, I and F. This follows from the fact that the commutation rules for the operators corresponding to J, I, F and the electric dipole moment D of the atom are the same as for the operators corresponding to L, S, J and D. Because these commutation rules determine the matrix elements of the electric dipole moment D, these elements must in the first case depend on the quantum numbers J, I and F in the same way as in the second case they depend on the quantum numbers L, S and J.

But from these dipole matrix elements follow immediately those for the electric quadrupole radiation. It seems therefore apparent that, as was first pointed out by Opęchowski (see Mrozowski 1940 a), we can use for the electric quadrupole radiation the same selection rules and intensity formulae in the case of hyperfine structure as in the case of fine structure, merely replacing the quantum numbers L, S, J by J, I, F. As Gerjuoy (1941) recognized, the same procedure is also applicable to the magnetic dipole radiation.

We can even expect that the expressions derived for the intensities under the assumption of the existence of the L, S and J vectors (LS coupling) will in the case of hyperfine structure be generally in better agreement with experiment than in the case of fine structure. The coupling between the total angular momentum J of the electrons of the atom and the angular momentum I of the nucleus will generally be weaker than that between the orbital and spin angular momenta (L and S). We can infer this immediately from the value of the corresponding splittings in the fine and hyperfine structures. J and I are, therefore, in the case of hyperfine structure, generally better quantum numbers than L and S in the case of fine structure.

The examination of the intensity distribution and the Zeeman effect in the hyperfine structure gives us, therefore, a new and a very reliable tool for the examination of the multipole character of a spectral line. Such investigations were carried out by Mrozowski (1940 a) first of all on the Hg II line $\lambda 2815$ ($^2D_{\frac{1}{2}} \rightarrow {}^2S_{\frac{1}{2}}$), which turned out to be a quadrupole line. He (Mrozowski 1940 b) performed also very careful photometric measurements in the Pb I spectrum on lines arising from transitions between the lowest levels of the electron configuration $6s^26p^2$. Because the Pb isotope 207, which is responsible for the hyperfine structure of the Pb spectrum, has the spin $\frac{1}{2}$, every fine structure level with $J \neq 0$ splits into two hyperfine structure levels with $F = J + \frac{1}{2}$ and $F = J - \frac{1}{2}$. Only the levels with $J = 0$ are not split up. In the case of the spectral line $\lambda 5313$ ($^1S_0 \rightarrow {}^3P_2$) the hyperfine structure is therefore given by a doublet. From the fact that the ratio of the intensities of both the components is $\frac{3}{7}$ we can infer already from the sum rule that we have here to do with an electric quadrupole radiation. The spectral line $\lambda 4618$ ($^1S_0 \rightarrow {}^3P_1$), however, is a doublet with intensity ratio $\frac{1}{2}$ and is therefore a magnetic dipole line. Finally, the line $\lambda 7330$ ($^1D_2 \rightarrow {}^3P_1$) was

shown to be a magnetic dipole line containing a small admixture of electric quadrupole radiation. Mrozowski (1946) also examined the hyperfine structure of the spectral lines of Bi I due to transitions between the lowest levels of the $6p^3$ electron configuration. All these are predominantly magnetic dipole lines.

The Zeeman effect of the hyperfine structure was investigated by Jenkins and Mrozowski (1941 a, b) for the magnetic dipole line $\lambda 4618$ and the electric quadrupole line $\lambda 5313$. The observed intensities were in satisfactory agreement with the theoretical predictions.

§ 6. THE TRANSITION PROBABILITIES

In § 3 we have discussed the experiments which enable us to establish the multipole character of the radiation of an atom by the investigation of the directional dependence of its radiation field. But such experiments can be carried out only on spectral lines emitted by sources of light in our laboratories and not on cosmic sources of light inaccessible for such investigations. Only one of the methods thus far described could be used in principle for this purpose, namely the investigation of the intensity distribution in the hyperfine structure (§ 5). We can hardly expect that this method will have astrophysical importance because of the Doppler breadths and low intensities of the lines in the spectra of cosmic objects.

The only quantities we can measure in cosmic spectra are line intensities and breadths. For the determination of the multipole character it is, however, only the natural breadth and not the observed Doppler breadth which is of theoretical importance. But from the natural breadth we could only infer the sum of the transition probabilities from the initial and final levels and so could only very indirectly draw conclusions about the multipole character of the line concerned.

Thus there is no other method for the determination of the multipole character of a cosmic line than that based on intensity measurements. For this purpose we have to measure the intensities of all lines in the multiplet which contains the line in question.

Such measurements must be compared with the theoretical transition probabilities. These are also necessary for a study of physical conditions in radiating gases in cosmic and laboratory sources. For such purposes as the determination of the population of different energy levels the transition probabilities for non-observable spectral lines of very low frequencies lying in the far infra-red are also important.

In this way we can understand the great efforts already made to calculate the relative and absolute transition probabilities. First of all the relative transition probabilities were calculated in the case of *LS*-coupling for the electric quadrupole (Rubinowicz 1930) and magnetic dipole radiations (Brinkman 1932). The next step was made by Stevenson (1932), who calculated the absolute transition probabilities for intermediate coupling cases for transitions between two levels belonging to the same electron configuration, i.e. p^2. But this author has determined only the sums of the transition probabilities of the second-class radiation. The calculation of the separate transition probabilities was carried out for the most prominent nebular lines by Condon (1934) for transitions within the electron configurations p^2, p^3 and p^4. At the same time Blaton (1934) calculated the transition probabilities for the magnetic dipole radiations within the p^2 configuration.

Pasternack (1940) re-examined these calculations and extended them to many more atoms also in the d^2 and d^3 configurations. Finally, Shortley (1940) has formulated the theory of the second-class multipole radiations in such a fashion as to make available all the methods of calculation and sum rules which have been developed for electric dipole radiation. He simplified especially the calculation of line strengths in the case of intermediate coupling by defining a square root of line strength in such a way that it transforms like a Hermitian matrix component. Using the results of these considerations, Shortley, Aller, Baker and Menzel (1941) have calculated the strengths of the magnetic dipole and electric quadrupole lines (in the latter case apart from the squares of integrals involving the radial wave functions) for the electron configurations p^2, p^3 and p^4 as functions of a single parameter χ embracing all cases of coupling from the pure *LS*- to the pure *jj*-coupling. Since careful estimates of probable χ values for the most important ions were given by Robinson (1937) and Robinson and Shortley (1937), or can be interpolated from the data of these authors, the problem of the calculation of the transition probabilities for transitions within the p^i configurations is completely solved by these investigations except for the radial integral mentioned above, which in atomic units is of the order of magnitude of unity.

§7. THE ELECTRIC OCTUPOLE AND MAGNETIC QUADRUPOLE RADIATIONS

The identification of octupole lines in the optical spectrum has been reported by Huff and Houston (1930) for the Hg I line $\lambda 2270$ ($^3P_2 \rightarrow {}^3S_0$) and by Jacquinot and Brochard (1943) for the He I lines $\lambda 4387{\cdot}21$ and $\lambda 4143{\cdot}43$ ($^1G_4 \rightarrow {}^1P_1$). But in both cases the authors themselves recognized that the lines in question have another origin. The Hg I line $\lambda 2270$ is caused by the nuclear spin of the odd Hg isotopes. The He I lines $\lambda 4387{\cdot}21$ and $\lambda 4143{\cdot}43$, however, which were identified by their transverse Zeeman effects, were found to be due to enforced radiation. In fact Brochard, Jacquinot and Pluvinage (1945) and Brochard and Jacquinot (1945) have shown that, under given circumstances, the transverse Zeeman effect of a spectral line enforced by the electric field of ions is the same as that for the radiation of the next higher order allowed for the transition in question, if there exists no component of the electric field parallel to the magnetic field causing the Zeeman effect (cf. also Yvon 1943 and Pluvinage 1947).

There are however many reasons why so far no multipole lines of the third class have been observed in the optical range.

(1) *The extremely low transition probabilities of the multipole radiation of the third class.* If we denote by a the radius of the radiating atom and by λ the wavelength, then the orders of magnitude of the transition probabilities of the electric octupole and magnetic quadrupole radiations are $(2\pi a/\lambda)^2$ times smaller than those of the electric quadrupole and magnetic dipole radiations respectively. Taking as a basis the estimates of Shortley *et al.* (1941) for the transition probabilities, viz. for electric dipole radiation $A_p\ 2.10^7\,\mathrm{sec}^{-1}$, for magnetic dipole radiation $A_m\ 3.10^2\,\mathrm{sec}^{-1}$, for electric quadrupole radiation $A_q\ 8.10^{-1}\,\mathrm{sec}^{-1}$, and taking $a = 0{\cdot}5$ A. and $\lambda = 4580$ A., we shall therefore expect the transition probabilities of the electric octupole and magnetic quadrupole radiations to be of the order of magnitude of $4.10^{-7}\,\mathrm{sec}^{-1}$ and $10^{-4}\,\mathrm{sec}^{-1}$ respectively. Since a day has about 10^5 sec. and a month about 2.10^6 sec., the time between two disturbances of the

excited atom (by collision with other atoms, electrons, photons, etc.) must be of the order of months or days if every excited atom is to lose its energy by radiation.

(2) *Masking of the multipole lines of the third class by electric dipole lines.* It is possible to reduce the selection rules for the multipole radiation of the third class to the selection rules for the multipole radiations of the first two classes. A transition between two energy levels A and B can take place as an electric octupole (or magnetic quadrupole) transition only when there exists a third level C such that an electric dipole transition can take place between one of the levels A or B and the level C and an electric quadrupole (or magnetic dipole) transition can take place between the other of the levels A or B and the level C. In this way we can establish the selection rules:

(*a*) A multipole radiation of the third class can take place like the electric dipole radiation only between two levels of opposite parity.

(*b*) For the octupole radiation transitions are allowed in which $|J-J'|\leqslant 3$ with exception of those for which $J+J'<3$ (additional prohibition). In the case of magnetic quadrupole radiation we have the selection rule $|J-J'|\leqslant 2$, the transitions $J+J'<2$ being excluded.

(*c*) In case of the LS coupling there are for L the analogous selection rules: for the electric octupole radiation $|L-L'|\leqslant 3$, excluded $L+L'<3$, and for the magnetic quadrupole radiation $|L-L'|\leqslant 2$, excluded $L+L'<1$.

(*d*) From the selection rule (*a*) it follows that the magnetic quadrupole transitions can take place only between two energy levels belonging to different electron configurations, because all levels belonging to a common electron configuration are of the same parity.

Moreover, we can state that in contrast with the case of magnetic dipole radiation the magnetic quadrupole radiation can be emitted in the visible also in the case of pure LS coupling, so that it need not be an intercombination radiation. This follows from the fact that with this coupling we can choose for example the states A and C as two states of the same multiplicity belonging to different electron configurations and C and B as two states of the same multiplicity belonging to the same electron configuration. Then between A and C an electric dipole radiation and between C and B a magnetic radiation can take place.

From these selection rules it follows that generally every electric dipole radiation is accompanied by a multipole radiation of the third class, excluding the cases where, for example, the latter radiation is forbidden by one of the additional prohibitions. Since we may expect that in the case of multipole radiation of the third class, as in the case of electric dipole radiation, those transitions are the most probable in which the changes of J and L are the same, the electric dipole radiation masks the strongest spectral lines in the third-class radiation multiplet. But there are of course also electric octupole multiplets which are not at all masked by electric dipole lines, for example where $2\leqslant|L-L'|\leqslant 3$.

(3) *Rareness of corresponding metastable states.* We can expect to observe the multipole radiation of the third class in nebulae only in the case where suitable metastable levels exist for this radiation, that is, levels from which there can take place no radiation of the first two classes to lower levels. For the appearance of the third-class multipole radiation the existence of such metastable levels is, however, not sufficient. There must exist, moreover, a lower level to which a third-class multipole transition from the metastable level can take place. But if such pairs of levels exist at all, they are extremely rare. Both levels must be of opposite

parity, and the electric dipole radiation must not be subject to the additional prohibitions ($J+J'<1$ or $L+L'<1$ in the case of LS couplings) since they would then exclude also the multipole radiation of the third class.

(4) Finally we have to mention the *importance of the additional prohibitions.* Their influence increases with the order of the multipole radiation. The greater this order the greater must be the sum of the J quantum numbers of the initial and final levels if a transition is to be possible. The additional prohibition excludes in a $(J-J')$ transition any multipole radiation of an order greater than $2^{J+J'}$. The spectral lines due to $(J-J')$ transitions contain therefore multipole radiations only up to the order $2^{J+J'}$. In a similar way in the case of LS coupling the additional prohibitions for the L quantum number come into play. These restrictions make themselves the more felt the lower the J and L values of both the levels.

Summarizing, we can say that we are very unlikely to find in the visible a multipole radiation of the third class.

Much more probable is the discovery of this type of radiation in the x-ray region. In the K or L series of uranium the intensities of the strongest octupole lines are about 10^{-4} of those of K_α or L_α. In fact some spectral lines there have been discussed in this region (e.g. $L_{II}-N_V$ in the Bi spectrum, Siegbahn 1931, p. 204) which, accepting the identification of terms, could be electric octupole lines. But against this identification we can quote x-ray lines which were attributed to transitions $K-L_I$ and L_I-N_I (Lindh 1930, p. 213) forbidden as spontaneous transitions for multipole radiations of all orders. The point is that these take place between two levels with the quantum number $J=\frac{1}{2}$ and of the same parity, so that we evidently have here enforced electric dipole transitions. But possibly the situation is similar to that in the case of the rare earths (§ 3). It could be that the spectral lines are due partly to enforced electric dipole and partly to spontaneous electric octupole transitions. With regard to the fact that the existence of the spectral lines under discussion, and still more their assignment to the transitions, is open to doubt, the first thing which must be done is a re-examination and extension of the experimental evidence.

Of a special importance are the multipole radiations of higher order in the case of nuclear γ-rays. The dipole moments of all but the lightest nuclei are reduced by the approximate coincidence of the centres of charge and mass. Therefore the electric dipole and quadrupole radiations are generally of comparable magnitude, but the remaining multipole radiations are weaker.

There are several independent experimental methods which have been devised for the determination of the multipole character of the γ-rays.

(1) *The measurement of the internal conversion coefficient* α, i.e. of the ratio of the probability of the emission of a conversion electron to the probability of the emission of a γ-quantum. Especially for low energy γ-rays α is very sensitive to the multipole character of the γ-radiation, but for higher energy γ-rays, 300 kev. or more, the change is not so rapid. Because the experimental determination of α is very difficult, this method is of no practical importance for higher energy γ-rays (Hebb and Uhlenbeck 1938, Dancoff and Morrison 1939, Hebb and Nelson 1940, Schafroth 1948).

(2) *The determination of the ratio* N_K/N_L *of the numbers of the K and L conversion electrons.* The value of N_K/N_L is very sensitive to the multipole order and is different for electric and magnetic radiations, so that it can be used to

distinguish these two types of multipole radiation. This method is especially noteworthy because N_K/N_L is easier to measure than the conversion coefficient (Hebb and Nelson 1940).

(3) *The life-time of γ-ray transitions.* If the transition takes place between two levels with the I quantum numbers I and I', the lowest possible order of the emitted electric or magnetic multipole radiation is $2^{|I-I'|}$ and $2^{|I-I'|-1}$ respectively. But if these transitions are forbidden by Laporte's rule the transitions of the next higher order have to take place. The process of internal conversion tends thereby to speed up the transitions, so that the lifetime of the transition calculated from electromagnetic considerations must be divided by $1+\alpha$, where α is the total conversion coefficient. These lifetimes are, however, observable only in cases where they are sufficiently high, i.e. if $|I-I'|$ is large enough (isomerism (Hebb and Uhlenbeck 1938)).

All these three methods have been used by Helmholz (1941) for the determination of the multipole character of nuclear γ-rays. In eight out of the fourteen γ-lines investigated a fairly definite assignment of the multipole order has been made. There were observed electric and magnetic multipole radiations up to the orders 2^5 and 2^4 respectively.

(4) *Directional correlation of successive γ-quanta* (Hamilton 1940). If a nucleus passes from an excited state A first to another excited state B and then to the ground state C emitting two successive γ-quanta there is some correlation between their directions, depending on the multipole character of the emitted radiations and of the angular momenta of the nuclear states involved. The reason may be stated briefly as follows: Let us first of all assume that the transition A→B starts from a definite magnetic substate of A. Then by the corresponding selection rules and transition probabilities are fixed the probabilities for the emission of γ-quanta in different directions in the transition A→B. But these transition probabilities determine the populations of the different substates of B, and so, by the selection rules and transition probabilities for the transition B→C, also the probabilities for the emission of γ-quanta in different directions in the second transition. Therefore the probabilities for the emission of γ-quanta in different directions are determined for a given substate of A and are thus correlated. This correlation persists obviously if we take the average value with respect to the initial substates of A. In these considerations we must, however, pay regard to the fact that the phases of the substates of B are related if the first γ-quantum has been emitted in a definite direction.

Goertzel (1946) has extended Hamilton's calculations to take into account the effect on the nucleus of the magnetic field generated by the orbital electrons. It is evident that the above correlation can be observed only if the disturbances caused by this magnetic field are sufficiently small during the lifetime of the nuclear state B.

Hamilton and Goertzel have considered only the cases of pure electric dipole and quadrupole and pure magnetic dipole transitions. But the electric quadrupole and magnetic dipole radiations can also be emitted simultaneously. In such a case the interference effect (cf. § 4) comes into play and changes the correlations (Ling and Falkoff 1948).

If in the radiation field of an electric multipole of the order 2^l we substitute H for E and $-E$ for H, we get the radiation field of a magnetic multipole of the same order. Because also the relative transition probabilities between magnetic

substates are the same for the electric and magnetic multipoles of the same orders it is clear that the directional correlation of two successive γ-quanta does not distinguish between a pure electric and a pure magnetic multipole radiation of the same order. But it is found to be possible to discriminate between both these radiations if one of the γ-quanta counters measures γ-quanta only in a definite polarization state (Hamilton 1947, 1948, Falkoff 1948).

Conclusive evidence for the existence of the directional correlation of successive γ-quanta was given by Brady and Deutsch (1947, 1948 a, b). They observed this effect in the case of the nuclei ^{60}Ni, ^{46}Ti, ^{24}Mg, ^{134}Ba, ^{88}Sr and ^{106}Pd using Kallmann scintillation counters. The directional correlation of the γ-radiations of the nuclei ^{60}Ni, ^{46}Ti, ^{24}Mg and ^{134}Ba are consistent with the emission of two electric quadrupole quanta and the nuclear momenta 4, 2, 0 for the three states A, B, C. In the case of ^{60}Ni and ^{46}Ti these spin assignments are in good agreement with considerations based on selection rules (Deutsch, Elliot and Roberts 1945). Any admixture of magnetic dipole radiation is here excluded as the changes of the angular momentum quantum numbers are given by 2. The correlation of the radiations of ^{88}Sr can be explained by an interference between the electric quadrupole and magnetic dipole radiations in accord with a disintegration scheme proposed by Peacock (1947) assuming that the angular momenta of the three states involved are 2, 1, 0. No satisfactory explanation has been given so far in the case of the directional correlation of the ^{106}Pd radiations. Assuming electric quadrupole radiations and the assignment 0, 2, 0 for the angular momentum quantum numbers to the three nuclear states concerned, an agreement between theory and experiment can be obtained only if the theoretical coefficients of the angular anisotropy are reduced by a factor of 2. But no reason could be found for such a procedure.

Deutsch and Metzger (1948) succeeded in observing the directional correlation in the case where one γ-quantum goes to a polarization-sensitive and the other to a polarization-insensitive detector. They used the γ-radiation of ^{106}Pd and found that the ratio of the counting rate of γ-quanta polarized in the plane of both the emitted quanta to the counting rate of γ-quanta polarized in the direction perpendicular to this plane corresponds to two electric quadrupole radiations of the ^{106}Pd nucleus, but only if the experimentally observed and not the theoretical coefficients of the directional correlation are used in the calculation.

(5) *Angular distribution of γ-quanta emitted following β-decay.* The probability that the direction of the emission of a γ-quantum emitted after a β-decay makes an angle ϕ with the direction of the β-ray depends upon the multipole character of the γ-radiation, the angular momenta of the three nuclear states involved and the β-interaction (Hamilton 1941). No experimental evidence of this effect has been found so far.

(6) *Directional correlation of successive conversion electrons.* We get a further group of correlation effects depending on the multipole character of the γ-radiation if we take into account the conversion electrons. Instead of two successive γ-quanta we can then have two successive conversion electrons or one conversion electron preceding or following a γ-quantum. The greater the multipole order of the γ-radiation the greater the probability of the emission of a conversion electron, but at the same time also the greater in general is the lifetime of the corresponding initial state. The importance of this group of phenomena therefore increases with increasing multipole order of the γ-radiation. But nevertheless

the multipole order of the second γ-radiation, emitted by the intermediary quantum state of the nucleus as initial level, must not be too high, because otherwise the lifetime of this state is too long, so that the disturbances caused by the electrons come into play and destroy the correlation effect. No calculations concerning these effects have been published so far. But recently, after this Report had been written, experiments were reported by Ward and Walker (1949) showing the directional correlation effect of two successive conversion electrons from ^{181}Ta.

The general form of the dependence of a directional correlation upon the angle between the directions in which photons and other particles involved are emitted can be derived using only the invariance properties of the physical process under space rotation and inversion (Yang 1948).

ACKNOWLEDGMENT

The author wishes to express his sincere thanks to the Editor for amendments in wording in this Report.

REFERENCES

ARAKI, G., 1948, *Phys. Rev.*, **74**, 472.
BETHE, H., 1929, *Ann. Phys., Lpz.*, **3**, 133.
BETHE, H. A., and SPEDDING, F. H., 1937, *Phys. Rev.*, **52**, 454.
BLATON, J., 1934, *Z. Phys.*, **89**, 155; 1937, *Acta Phys. Pol.*, **6**, 256.
BLATON, J., and NIEWODNICZAŃSKI, H., 1934, *Phys. Rev.*, **45**, 64.
BLEANEY, B., 1946, *Physica*, **12**, 595.
BOWEN, I. S., 1936, *Rev. Mod. Phys.*, **8**, 55.
BOWEN, I. S., and SWINGS, P., 1947, *Astrophys. J.*, **105**, 407.
BRADY, L., and DEUTSCH, M., 1947, *Phys. Rev.*, **72**, 870; 1948 a, *Ibid.*, **73**, 1266; 1948 b, *Ibid.*, **74**, 1541.
BRINKMAN, H. C., 1932, *Zur Quantenmechanik der Multipolstrahlung* (Groningen: Noordhoff).
BROCHARD, J., and JACQUINOT, P., 1945, *Ann. Phys. Paris*, **20**, 508.
BROCHARD, J., JACQUINOT, P., and PLUVINAGE, P., 1945, *C.R. Acad. Sci., Paris*, **220**, 38.
BROER, L. J. F., 1946, *Physica*, **12**, 642.
BROER, L. J. F., GORTER, C. J., and HOOGSCHAGEN, J., 1945, *Physica*, **11**, 231.
CONDON, E. U., 1934, *Astrophys. J.*, **79**, 217.
DANCOFF, M., and MORRISON, P., 1939, *Phys. Rev.*, **55**, 122.
DEUTSCH, M., ELLIOT, L. G., and ROBERTS, A., 1945, *Phys. Rev.*, **68**, 193.
DEUTSCH, M., and METZGER, F., 1948, *Phys. Rev.*, **74**, 1542.
DEUTSCHBEIN, O., 1940, *Ann. Phys., Lpz.*, **36**, 183.
DOERMANN, F. W., 1938, *Phys. Rev.*, **53**, 420.
DOERMANN, F. W., and HALPERN, O., 1939, *Phys. Rev.*, **55**, 486.
DURAND, E., 1946 a, *C.R. Acad. Sci., Paris*, **222**, 1044; 1946 b, *J. Phys. Radium*, **7**, 363.
EDLÉN, B., 1942, *Z. Astrophys.*, **22**, 30; 1945, *Mon. Not. R. Astr. Soc.*, **105**, 323.
FALKOFF, D. L., 1948, *Phys. Rev.*, **73**, 518.
FREED, S., 1942, *Rev. Mod. Phys.*, **14**, 105.
FREED, S., and WEISSMAN, S. J., 1941, *Phys. Rev.*, **60**, 440.
FRERICHS, R., and CAMPBELL, J. S., 1930, *Phys. Rev.*, **36**, 1460.
GERJUOY, E., 1941, *Phys. Rev.*, **60**, 233.
GOBRECHT, H., 1937, *Ann. Phys., Lpz.*, **28**, 673.
GOERTZEL, G., 1946, *Phys. Rev.*, **70**, 877.
GORDY, W., 1948, *Rev. Mod. Phys.*, **20**, 668.
HALPERN, O., and DOERMANN, F. W., 1937, *Phys. Rev.*, **52**, 937; 1939, *Ibid.*, **55**, 681.
HAMILTON, D. R., 1940, *Phys. Rev.*, **58**, 122; 1941, *Ibid.*, **60**, 168; 1947, *Astrophys. J.*, **106**, 457; 1948, *Phys. Rev.*, **74**, 782.
HANSEN, W. W., 1935, *Phys. Rev.*, **47**, 139.
HEBB, M. H., and NELSON, E., 1940, *Phys. Rev.*, **58**, 486.
HEBB, M. H., and UHLENBECK, G. E., 1938, *Physica*, **5**, 605.

HEITLER, W., 1936, *Proc. Camb. Phil. Soc.*, **32,** 112.
HELLWEGE, K.-H., 1941, *Z. Phys.*, **117,** 198; 1942, *Ibid.*, **119,** 325; 1943, *Ibid.*, **121,** 588.
HELMHOLZ, A. C., 1941, *Phys. Rev.*, **60,** 415.
HUFF, L. D., and HOUSTON, W. V., 1930, *Phys. Rev.*, **36,** 842.
HUMBLET, J., 1943, *Physica*, **10,** 585; 1944 a, *Ibid.*, **11,** 91; 1944 b, *Ibid.*, **11,** 100; 1946, *C.R. Acad. Sci. Paris*, **223,** 419.
JACQUINOT, P., and BROCHARD, P., 1943, *C.R. Acad. Sci., Paris*, **216,** 581.
JENKINS, F. A., and MROZOWSKI, S., 1941 a, *Phys. Rev.*, **59,** 808; 1941 b, *Ibid.*, **60,** 225.
KELLOGG, B. M., and MILLMAN, S., 1946, *Rev. Mod. Phys.*, **18,** 323.
KRAMERS, H. A., 1938, *Die Grundlagen der Quantentheorie, Quantentheorie des Elektrons und der Strahlung* (Leipzig : Akad. Verlags.), p. 419; 1943, *Physica*, **10,** 261.
LAUE, M., 1941, *Röntgenstrahlinterferenzen* (Leipzig : Akad. Verlags.).
LINDH, A. E., 1930, *Röntgenspektroscopie, Handb. Exp. Phys.*, **24,** Pt. 2.
LING, D. S., and FALKOFF, D. L., 1948, *Phys. Rev.*, **74,** 1224.
MILIANCZUK, B., 1935, *Bull. Acad. Pol. Sci. et Lettr. Classe de Sci. Math. et Nat.*, Série A, 430.
MROZOWSKI, S., 1940 a, *Phys. Rev.*, **57,** 207; 1940 b, *Ibid.*, **58,** 1086; 1944, *Rev. Mod. Phys.*, **16,** 153; 1946, *Phys. Rev.*, **69,** 169.
NIEWODNICZANSKI, H., 1933, *Phys. Rev.*, **44,** 854; 1934 a, *Acta Phys. Pol.*, **2,** 375; 1934 b, *Ibid.*, **3,** 285; 1936, *Ibid.*, **5,** 111.
PASTERNACK, S., 1940, *Astrophys. J.*, **92,** 129.
PEACOCK, W. C., 1947, *Phys. Rev.*, **72,** 1049.
PLUVINAGE, P., 1947, *Ann. Phys., Paris*, **2,** 486.
ROBERTS, A., BEERS, Y., and HILL, A. G., 1946, *Phys. Rev.*, **70,** 112.
ROBINSON, H. A., 1937, *Phys. Rev.*, **52,** 724.
ROBINSON, H. A., and SHORTLEY, G. H., 1937, *Phys. Rev.*, **52,** 713.
ROSA, A. M., 1943, *Ann. Phys., Lpz.*, **43,** 161.
RUBINOWICZ, A., 1930, *Z. Phys.*, **65,** 662.
RUBINOWICZ, A., and BLATON, J., 1932, *Ergebn. exakt. Naturwiss.*, **11,** 176.
SCHAFROTH, R., 1948, *Helv. Phys. Acta*, **21,** 499.
SEGRÈ, E., 1931, *Nouvo Cim.*, **8,** No. 2.
SEGRÈ, E., and BAKKER, C. J., 1931, *Z. Phys.*, **72,** 724.
SELÉNYI, P., 1911, *Ann. Phys., Lpz.*, **35,** 444; 1939, *Phys. Rev.*, **56,** 477.
SHORTLEY, G. H., 1940, *Phys. Rev.*, **57,** 225.
SHORTLEY, G. H., ALLER, L. H., BAKER, J. G., and MENZEL, D. H., 1941, *Astrophys. J.*, **93,** 178.
SIEGBAHN, M., 1931, *Spektroscopie der Röntgenstrahlen*, 2nd ed. (Berlin : Springer).
SPEDDING, F. H., 1940, *Phys. Rev.*, **58,** 255.
STEVENSON, A. F., 1932, *Proc. Roy. Soc.* A, **137,** 298.
SWINGS, P., 1945, *Pub. A.S.P.*, **57,** 117.
VAN VLECK, J. H., 1937, *J. Phys. Chem.*, **41,** 67.
WARD, A. H., and WALKER, D., 1949, *Nature, Lond.*, **163,** 168.
YANG, C. N., 1948, *Phys. Rev.*, **74,** 764.
YVON, J., 1943, *C.R. Acad. Sci., Paris*, **216,** 585, 852.

COLLISIONS BETWEEN ATOMS AND MOLECULES AT ORDINARY TEMPERATURES

By H. S. W. MASSEY
University College, London

CONTENTS

PAGE
§1. Introduction 248
§2. Elastic collisions 249
2.1. The effective cross-sections 249
2.2. The interaction between helium atoms 251
2.3. Measurement of total collision cross-sections by the molecular ray method . 254
2.4. The mobility of positive ions in gases 256
§3. Collisions involving change of internal energy 258
3.1. Inelastic collisions between atoms—exchange between electronic and translational energy 258
3.2. Inelastic collisions between atoms and molecules—exchange between electronic, vibrational and translational energy 262
3.3. Inelastic collisions between atoms and molecules—exchange between vibrational and translational energy 264
3.4. Inelastic collisions between atoms and molecules—exchange between rotational and translational energy 267

§1. INTRODUCTION

THE experimental study of the phenomena associated with the slow collisions of atoms and molecules must usually be carried out under conditions in which the relative velocity involved in the impacts is not definite but distributed according to the usual Maxwell law about a mean determined by the temperature. Nevertheless, a growing body of information is now available about these collisions and in some cases a satisfactory theory has been developed to interpret and supplement the observed data. In this review we shall confine ourselves to conditions corresponding to ordinary temperatures and also somewhat arbitrarily exclude any but incidental discussion of chemical reaction rates.

We shall first discuss elastic collisions. These determine the coefficient of viscosity, and the thermal conductivity of a gas, as well as the coefficient of diffusion of one gas in another or of ions through a gas (ionic mobility). It is from an analysis of the magnitude and temperature dependence of these coefficients that much information has been derived about the interactions between atoms. The quantum theory has shown that a further quantity, which is capable of direct measurement, and may also be used for this purpose, is the total collision cross-section for impacts between the atoms. §2 will be devoted to a review of these various aspects of elastic collisions.

In collisions between molecular systems, one of which possesses excess energy of internal motion, it is of interest to discover the probability that, on impact with a normal molecule, the energy be redistributed in other degrees of freedom. The problem is essentially that of determining the extent to which energy transfer between relative translation, electronic motion, vibration and rotation takes place. This is of considerable importance in such phenomena as the quenching of fluorescence, sensitized fluorescence, the dispersion of sound and the rate of unimolecular reactions. §3 will include a brief review of these transfer collisions and their importance in certain branches of physics.

§ 2. ELASTIC COLLISIONS

2.1. *The Effective Cross-Sections*

It is convenient to discuss first some of the features of the quantum theory of collisions. We shall throughout refer the motion of the two colliding atoms or molecules to the centre of mass as origin, so that θ, the angle of scattering in this system, is the angle through which the direction of relative motion is turned in the impact. The differential cross-section is such that the number of collisions which occur per cm^3 per sec., between two kinds of atoms which are moving with relative velocity v, in which the direction of relative motion is turned through an angle between θ and $\theta+d\theta$, is

$$2\pi N_1 N_2 v I(\theta) \sin\theta \, d\theta, \qquad \ldots\ldots(1)$$

N_1, N_2 being the number of each kind of atoms present per cm^3. The total elastic collision cross-section Q_0 is given by

$$Q_0 = 2\pi \int_0^\pi I(\theta) \sin\theta \, d\theta. \qquad \ldots\ldots(2)$$

It is such that the number of collisions taking place can be calculated by treating the atoms as rigid spheres of radius $(Q_0/4\pi)^{\frac{1}{2}}$.

According to the classical theory Q_0, defined in this way, would not be finite if the interaction between the atoms falls gradually to zero and has no sharp boundary. The deviation produced by even a weak interaction at very great distances would be finite and the effective sizes of the atoms unbounded. In quantum theory, owing to the uncertainty principle, no meaning can be attached to deviations through less than a certain angle—they are unobservable. As a result of this the total cross-section Q_0 defined by (2) is finite in the quantum theory, provided the interaction between the atoms falls off more rapidly, at large distances, than the square of the inverse distance between them (Mott 1930, Mott and Massey 1949a).

If means can be found for measuring Q_0 another useful quantity for providing information about interatomic forces is available. It is important to notice that Q_0 can never be calculated by classical theory as the contribution from very small angles of scattering is always given incorrectly by that theory.*

A remarkable example of the fact that the classical theory can never give $I(\theta)$ correctly for all θ, except for the very special case of the Coulomb field, is provided by the problem of the rigid sphere. If the atoms are regarded as rigid spheres of diameter a, with no other interaction, the classical collision cross-section is, of course, finite and equal to πa^2. When the wavelength λ of the relative motion is large compared with a, the quantum theory gives a different value, $4\pi a^2$, which is not surprising as the wavelength is large compared with the size of the obstacle. But what is rather more unexpected is that the quantum theory, in the limit of very short wavelengths, gives, not a cross-section πa^2, but one of twice that value. The classical form for $I(\theta)$ is a constant, $a^2/4$. The quantum theory gives the same value of $I(\theta)$ for $\theta \gg \lambda/2\pi a$, but for smaller values of θ it predicts that $I(\theta)$ should increase rapidly to reach $\pi^2 a^4/\lambda^2$ at $\theta = 0$ (see the schematic representation, Figure 1). It is the additional contribution from this

* It is possible to evaluate the cross-section $Q(\theta_0)$ for scattering through angles greater than θ_0 by classical theory provided that theory is valid for calculating the scattering through angles as small as θ_0. This may be made use of in practice by ensuring that deviations through angles less than θ_0 will not be observed.

region which doubles the classical cross-section. The effect is an example of edge diffraction and it has been discussed by Massey and Mohr (1933) and by Wergeland (1945).

The viscosity and heat conductivity of a gas depends on the ease of transport of momentum from one part of a gas, in which the mean kinetic energy is higher, to another. The more frequent the collisions the less easy the transport, but collisions in which the energy of the colliding partners is evenly shared after impact are more effective in this regard than others. The collision cross-section Q_η effective in viscosity is given by

$$Q_\eta = 2\pi \int_0^\pi I(\theta) \sin^3\theta \, d\theta, \qquad \ldots\ldots(3)$$

greater weight being given to deviations through 90°. The complete formula for the viscosity coefficient η of a gas has been given by Chapman (1916, 1917) and by Enskog (1917). It is

$$\eta = \frac{10}{\boldsymbol{k}^3 T^3 M^2} \left(\frac{4\pi \boldsymbol{k} T}{M}\right)^{3/2} \frac{1+\epsilon}{\pi R_{11}}, \qquad \ldots\ldots(4)$$

where M is the mass of a gas atom, T the absolute temperature, $\boldsymbol{k}$ Boltzmann's constant and

$$R_{11} = \tfrac{1}{2} \int_{-\infty}^{\infty} v^7 Q_\eta(v) \exp\{-Mv^2/4\boldsymbol{k}T\} \, dv. \qquad \ldots\ldots(5)$$

ϵ is a small correction, never greater than 0·017.

A further effective cross-section Q_D is required for the theory of diffusion. This is defined by

$$Q_D = 2\pi \int_0^\pi (1-\cos\theta) I(\theta) \sin\theta \, d\theta. \qquad \ldots\ldots(6)$$

Diffusion is retarded more effectively by those collisions in which $\theta \simeq 180°$ so that the paths of the colliding atoms are nearly reversed in sense—hence the additional weighting at these angles. The coefficient of diffusion D for two gases is given by (Chapman 1916, 1917, Enskog 1917)

$$D = \frac{3}{16} \pi^{\frac{1}{2}} \left\{2 \frac{(M_1+M_2)}{M_1 M_2} \boldsymbol{k}T\right\}^{7/2} \frac{1}{(n_1+n_2) P_{12}} \frac{1}{1-\epsilon_0}, \qquad \ldots\ldots(7)$$

where

$$P_{12} = \int_{-\infty}^{\infty} v^5 Q_D(v) \exp\{-\tfrac{1}{2} M_1 M_2 v^2 / \boldsymbol{k}T(M_1+M_2)\} \, dv. \qquad \ldots\ldots(8)$$

M_1, M_2 are the masses of the atoms of the respective gases and n_1, n_2 are the concentrations in number of atoms/cm^3. ϵ_0 is again an unimportant correction, never greater than 0·136.

The mobility of an ion in a pure gas is given by the relation

$$K = eD/\boldsymbol{k}T, \qquad \ldots\ldots(9)$$

where D is given by the formula (7).

Unlike the total cross-section the viscosity and diffusion cross-sections may usually be calculated with accuracy by classical methods. Of all monatomic

gases, helium is the only one for which quantum theory must be used to analyse, in terms of atomic interactions, the viscosity at temperatures above the liquefying point. At temperatures below about 80° K. the classical theory becomes very inaccurate for this gas. A discussion of the quantum modifications for rigid spherical atoms has been given by Massey and Mohr (1933).

We now discuss certain applications of these formulae.

2.2. *The Interaction between Helium Atoms*

An accurate determination of the interaction energy between helium atoms at various separations is important for several reasons. In particular, a satisfactory theory of the peculiar behaviour of liquid helium may well depend on the detailed nature of this interaction. Evidence is available from three sources, measurements of the viscosity and second coefficient over a wide range of temperatures and study of the scattering of helium atom beams by helium. The information obtained is still incomplete and not very consistent.

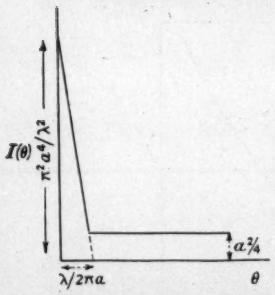


Figure 1. Schematic form of the relative angular distribution function $I(\theta)$ for collisions of rigid spheres of diameter a, λ being the de Broglie wavelength of the relative motion.

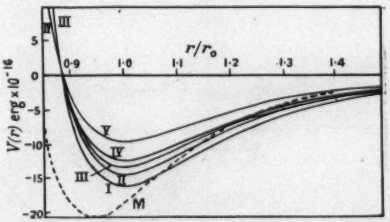


Figure 2. Illustrating various assumed potentials $V(r)$ for interaction between helium atoms plotted as functions of r/r_0 where $r_0 = 2{\cdot}945$ A.

At large separation there is a weak van der Waals attraction which changes rapidly to a strong repulsion. The depth of the attractive potential is so small that it is only from low temperature data that information can be derived about this part of the interaction.

2.21. *Evidence from low temperature experiments.* There exist seven determinations of the viscosity (van Itterbeek and Keesom 1938), and thirteen of the second virial coefficient (Holborn and Otto 1926, Nijhoff, Iliin and Keesom 1928, Keesom and Kraak 1935, Keesom and Walstra 1939), of helium at temperatures below 30° K.

The first analysis of the viscosity data using quantum theory to calculate the viscosity cross-sections was made by Massey and Mohr (1933, 1934). They assumed the interaction to be of the form given by Slater (1928), viz.

$$V(r) = be^{-ar} - cr^{-6}, \qquad \ldots\ldots(10)$$

where $b = 7{\cdot}7 \times 10^{-10}$ erg, $a = 4{\cdot}6 \times 10^{8}\,\mathrm{cm}^{-1}$ and $c = 1.37 \times 10^{-60}\,\mathrm{erg\,cm}^{6}$, and

found fair agreement with the observations. Since then a number of attempts have been made to improve on this potential, taking into account the second virial coefficient data also, the quantum formula for this coefficient having been derived by Gropper (1937) and by Beth and Uhlenbeck (1937). The most recent analysis is that of Buckingham, Hamilton and Massey (1941). They considered the family of interactions illustrated in Figure 2 in all of which the zero occurs at 2·945 A. The asymptotic form of these interactions was given by $-c/r^6$ but with different constants c for each. Curve IV corresponds to the Slater potential (10), curve I to one of the same form but 30% stronger, and curve III to a potential assumed by de Boer and Michels (1939) in their calculations. Of these interactions I and II give rise to no discrete stationary state whereas IV and V give one such state with binding energy 6×10^{-6} and $1{\cdot}5\times10^{-8}$ ev. respectively. It is not certain whether interaction III does or does not give a bound state but if so it must have extremely small binding energy.

The evidence from the virial coefficient data is illustrated in Figure 3. The

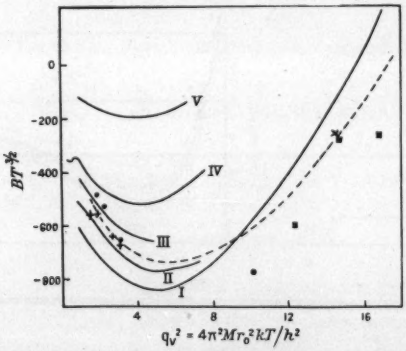


Figure 3. Comparison of observed and calculated virial coefficients (B) plotted as functions of $q_0^2=4\pi^2Mr_0^2kT/h^2$ where M is the mass of a helium atom and $r_0=2{\cdot}945$ A.
Full curves, calculated results from corresponding assumed potentials I–V illustrated in Figure 2.
– – – – – – – – Keyes' smoothing function.
Experimental points: × Holborn and Otto (1926).
■ Nijhoff, Iliin and Keesom (1928).
● Keesom and Kraak (1935).
+ Keesom and Walstra (1939).

data do not lie on a smooth curve and a considerable gap exists between 5° and 15°.

Keyes * has constructed an empirical function which provides a representation of the data between 2° K. and 650° K. which gives a good fit to the experimental points over the full temperature range except near 15° K. The virial coefficients calculated by use of the interactions I–IV are illustrated in the curves I–IV of Figure 3. It will be seen that the best fit is likely to be obtained by an interaction falling between II and III of Figure 2.

* Private communication to Dr. R. A. Buckingham.

This conclusion is not consistent with the viscosity data. The observed viscosity between 1·64° K. and 5° K. is fitted best by interaction V of Figure 2 and, if anything, suggests an interaction slightly weaker still. On the other hand, at temperatures between 15° and 30° K., the best fit is with the interaction I of Figure 2.

It is not clear whether the lack of consistent agreement is due to imperfection in the forms of the assumed interaction or to inaccuracy in the experimental data. Although the calculations were carried out with a fixed value of the distance r_0 at which the zero occurs it is unlikely that any adjustment of this distance will eliminate the inconsistencies. There is clearly a need here for further experimental work*.

In carrying out the calculations by the quantum theory allowance is made for the fact that helium atoms obey the Bose–Einstein statistics. The importance of this may be judged from the calculations carried out by Massey and Buckingham (1938) with the field I of Figure 2. They compared the viscosity cross-sections Q_η obtained on the respective assumptions of Bose–Einstein, classical and Fermi–Dirac statistics. Their results are illustrated in Figure 4 and it will be seen that

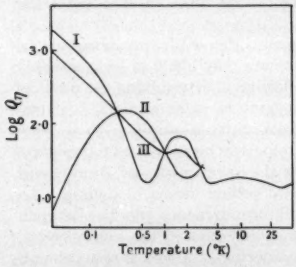


Figure 4. Calculated viscosity cross-sections Q_η for helium using the interaction *I* of Figure 2. The relative velocity v of the atom is given in terms of a temperature T where $v=3(kT/M)^{\frac{1}{2}}$, M being the mass of a helium atom. v is thus the most important relative velocity in determining the viscosity at temperature T.

Curve I—assuming Bose–Einstein statistics.
Curve II—assuming Fermi–Dirac statistics.
Curve III—assuming classical (Boltzmann) statistics.

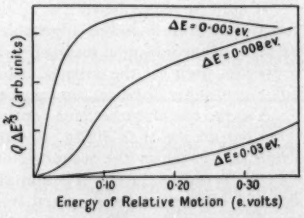


Figure 5. Illustrating the variation of the cross-section for transfer of excitation between atoms as a function of the energy of relative motion, for three energy discrepancies ΔE.

it is only at temperatures below 3° K. that the nature of the statistics is at all important.

2.22. *The evidence from medium and high temperatures.* The evidence from the measured viscosity of helium in the temperature range from 20° to 200° K. has been discussed most recently by Amdur (1947). He calculated the viscosity from the formulae (3), (4) and (5) using classical theory and the interaction energy proposed by Margenau (1939). This is given by

$$V(r)=[7{\cdot}70\,e^{-4{\cdot}60r}-5{\cdot}60\,e^{-5{\cdot}33r}-0{\cdot}0139r^{-6}-0{\cdot}030r^{-8}-0{\cdot}035r^{-10}]\times 10^{-10}\text{ ergs},$$

* Some clarification is provided from measurements by Ubbink and de Haas (1943) of the thermal conductivity of helium gas at 1·5 and 3° K. Their results are consistent with an interaction between II and III (Figure 2), agreeing with the evidence from the second virial coefficient.

where r is in A., and is illustrated by curve M of Figure 2. Amdur obtained agreement within 5% for the viscosity over the whole temperature range considered. This leaves the situation rather more confused than ever because the low temperature evidence does not appear to be consistent with Margenau's interaction. This may easily be seen by reference to Figure 2, which shows that this interaction includes a considerably stronger long range attraction than any used in the calculations of Buckingham, Hamilton and Massey. Thus the minimum is rather deeper and the zero occurs at a smaller distance. Furthermore, the analysis of the second virial coefficient of helium in the temperature range 70–673° K., carried out by Buckingham (1938), does not support the Margenau potential. Thus he concludes that the zero occurs for $r=2{\cdot}65$ A. and finds the data consistent with the interaction

$$V(r)=[8{\cdot}71e^{-4{\cdot}68r}-0{\cdot}0147r^{-6}]\times 10^{-10}\text{ ergs.} \qquad \ldots\ldots(11)$$

This interaction gives a slightly harder repulsion than that,

$$V(r)=[4{\cdot}39r^{-12}-0{\cdot}0152r^{-6}]\times 10^{-10}\text{ ergs,}$$

derived by Hirschfelder, Ewell and Roebuck (1938) from an analysis of the Joule–Thomson coefficients of helium.

It must be remembered that all the interactions derived by assuming special functional forms including adjustable constants are only likely to be reasonably accurate over the limited range of r which is effective in determining the physical properties used for the analysis, over the temperature range available. Extrapolation to other distances can only be carried out with caution.

A severe test of interactions derived from properties measured up to temperatures not greater than 700° K. is provided by the experiments of Amdur and Pearlman (1941) on the scattering of energetic helium atoms in helium. In these experiments, beams of helium atoms, with homogeneous energies between 300 and 1,000 ev., were prepared from helium ion beams of the same energy, by charge transfer. The cross-section for scattering of these atoms through angles greater than 4·5° in collision with helium atoms at room temperature was measured by observing the absorption of the beam in helium gas. By comparing these observations with calculations based on Buckingham's potential (equation (11)) it was found that this potential gives rather too large a repulsion for $r<0{\cdot}8$ A. but, even at 0·55 A., it is only about twice too large. This indicates a wider range of validity of the potential than would be expected.

2.3. *Measurement of Total Collision Cross-Sections by the Molecular Ray Method*

The measurement of the total collision cross-section Q_0 by molecular ray technique is simple in principle. Thus to measure Q_0 for two kinds of atoms A and B, averaged over the distribution of relative velocities fixed by the temperature of the gas, a narrow molecular beam composed of the atoms A is fired through the gas of atoms B contained in a small scattering chamber. The intensity of the beam is measured after passage through this chamber, both when the gas B is present at a measured pressure and when it is absent. From the ratio of these two intensities the mean cross-section may be determined, using the standard formulae of the kinetic theory of gases.

In order that a definite cross-section be obtained which is effectively independent of the angular resolving power of the apparatus it is necessary that this resolving power be very high. As a rough guide for testing whether any parti-

cular design is adequate a schematic angular distribution function $I(\theta)$, illustrated in Figure 1, may be used. This is based on the analogy with the rigid sphere problem. If a is the sum of the gas kinetic radii of the atoms, estimated from viscosity measurements or in some other way, then $I(\theta)$ may be taken as equal to $a^2/4$ for $\theta > \lambda/2\pi a$ and to rise linearly from this value to $\pi^2 a^4/\lambda^2$ at $\theta = 0$, λ being the wavelength of mean relative motion of the atoms. The apparatus should be adequate for the purpose if, with this distribution, the ratio

$$\int_{\theta_0}^{\pi} I(\theta) \sin\theta \, d\theta \Big/ \int_{0}^{\pi} I(\theta) \sin\theta \, d\theta, \qquad \ldots\ldots(12)$$

is sufficiently close to unity, θ_0 being the smallest angular deviation which can be detected by the apparatus.

Almost all experimental work in this direction has been carried out using molecular beams of alkali metal atoms. This is because the intensity of such a beam may be measured very conveniently and accurately by the hot filament method. The atoms undergo surface ionization and the current of ions produced is easily measured.

The earliest work on the collisions of atoms was that of Broadway (1933). It was directed to establishing the validity of the quantum result that $\mathrm{Lim}_{\theta \to 0} I(\theta)$ is finite. He observed the scattering, through angles between 0·2° and 1°, of a beam of sodium atoms by a beam of mercury atoms. Evidence in favour of the finite limit was found.

Two detailed experimental measurements of the cross-sections of the alkali metal atoms for collision with various gas atoms have been carried out, the first and most extensive by Rosin and Rabi (1935) and the most recent by Estermann, Foner and Stern (1947). The resolving power in the latter experiments was very high, the angle θ_0 of (12) being about 5″. In Rosin and Rabi's apparatus it was about 1′.

It was an interesting feature of Rosin and Rabi's experiments that the distribution of intensity across the beam remained the same in form with and without gas present in the scattering chamber. This absence of broadening by the scattering gas indicated that the probability of deviations taking place through angles less than the resolving power of the apparatus was small.

Rosin and Rabi measured cross-sections for collisions of each of the alkali metal atoms Li, Na, K, Rb and Cs with each of the rare gas atoms He, Ne and A. Their results are given in Table 1. The only measurements of Estermann, Foner and Stern which can be compared are those for Cs and He collisions. They find a cross-section 2½ times greater than that of Rosin and Rabi (see Table 1)

Table 1. Observed Collision Cross-Sections for Alkali metal–Rare gas Atom Impacts and the derived van der Waals Constants

Alkali metal atom	He			Ne			A		
	Q	$C(1)$	$C(2)$	Q	$C(1)$	$C(2)$	Q	$C(1)$	$C(2)$
Li	106	14·4	17	120	18·7	32	303	188	125
Na	130	17·8	(26)	213	40·4	(51)	401	192	(200)
K	165	31·7	35	259	50·5	68	580	356	260
Rb	152	25·4	(40)	268	49·0	(77)	572	249	(290)
Cs	162	29·4	44	287	56·7	87	572	235	325

Q, cross section in 10^{-16} cm².
C, derived van der Waals' constant in 10^{-60} cm⁶ : (1) from observed Q, (2) from polarizabilities.

and it is not clear how the discrepancy arises. The resolving power used by the latter authors should have been adequate.

A theoretical analysis of these results (Massey and Buckingham 1936), may be made on the plausible assumption that the van der Waals attraction, of the form $-c/r^6$, where r is the distance between the atoms and c a certain constant, is responsible for the scattering and is so large that the effect of the repulsion at short distances is quite negligible. For these conditions Massey and Mohr (1934) derived the formula

$$Q_0 = \frac{9}{4}\pi\left(\frac{3}{32}C\lambda\right)^{2/5},$$

where λ is the wavelength of the relative motion. Using this formula and a suitable mean for λ, values of the constant C for different pairs of interacting atoms may be derived. These are given in Table 1. It is further possible to estimate C if the polarizabilities of the atoms are known (Hellmann 1935). Those of the rare gas atoms are available from their refractivities (C. and M. Cuthbertson 1911) and those of lithium, potassium and caesium have been measured directly by Scheffers and Stark (1934) using molecular ray methods. The estimated values of C obtained in this way are also given in Table 1, interpolated values being included for sodium and rubidium. It will be seen that, on the whole, the agreement between estimated and derived values of C is not unsatisfactory. There is a tendency throughout for the derived values to be too small for rubidium and caesium. On the other hand, the value derived for the Cs–He case from the measurements of Estermann, Foner and Stern is many times greater than the estimated.

It seems clear that the measurement of total collision by molecular ray methods is likely to provide a powerful tool for investigating the interactions between atoms but much more experimental work is still required to establish it on a reliable foundation.

2.4. *The Mobility of Positive Ions in Gases*

A considerable amount of experimental work has been carried out on the mobilities of positive ions in gases. It has been described by Tyndall (1938) in his book *The Mobility of Positive Ions in Gases.*

For most cases a classical theory is adequate. The interaction energy between the ion and atom may be represented as of the rigid sphere form with a superposed attraction due to the polarization of the atom by the ion. Thus the interaction when the ion and atom are at a distance r apart is taken to be

$$V(r) = -\tfrac{1}{2}\alpha e^2/r^4, \quad r > a; \qquad V(r) \to \infty, \quad r < a,$$

where α is the polarizability of the atom. The radius a may be obtained from gas kinetic data and α from refractivity data or by molecular ray experiment (see §2.3). Alternatively the interaction may be taken to be of the form $V(r) = ar^{-n} - br^{-6} - \frac{1}{2}\alpha e^2/r^4$, where n is a positive integer greater than 6, the value 10 being mathematically most convenient. The first two terms are taken to be the same as for interaction between the neutral atoms, the form being that assumed by Lennard-Jones (1931) in his study of the equation of state of monatomic gases. Given $V(r)$, $I(\theta)$ may be calculated by classical methods, and hence the mobility by means of formulae (6), (7), (8) and (9). Calculations carried out in these ways by Hassé and Cook (1931) have been found to give satisfactory agreement with experiment for a number of cases.

When considering the mobility of an ion such as He^+ in a gas of the same kind it is necessary to allow for the fact that charge transfer may occur on collision: $He^+ + He \rightarrow He + He^+$. Because of this, it is not a good approximation to calculate the mobility assuming that the interaction between, say, He^+ and He is given by adding a polarization potential to that between two helium atoms. The correct procedure was given by Massey and Smith (1933) and applied by Massey and Mohr (1934) to calculate the mobility of He^+ in He.

According to quantum mechanics there are two possible interactions between a helium ion and atom in their ground states, depending on whether the electron distribution is symmetrical or antisymmetrical with respect to the two nuclei. Corresponding to each interaction there is associated an amplitude for scattering between angles θ and $\theta + d\theta$. The correct amplitude to take is a suitable linear combination of these two. The symmetrical interaction includes a powerful long range attraction which makes such a large contribution to the scattered amplitude that the effective diffusion cross-section is considerably larger than would be the case if charge exchange were ignored.

The mobility calculated in this way for He^+ in He was found to be 12 cm^2/sec. volt. This is in marked disagreement with the value 21·4 cm^2/sec. volt, for the ions from a helium glow discharge, observed by Tyndall and Powell (1931). The reason for this discrepancy is still not clear. Meyerott (1944) has suggested that the ions concerned in the experiments were not He^+ but He_2^+. He points out further that, in earlier measurements by the same authors (Tyndall and Powell 1930) in which an α-particle source of ions was used, the value 13cm^2/sec. volt was found, in much closer agreement with the calculated value. On the other hand, Mott (private communication) has drawn attention to the possibility that the neglect of certain terms in the calculation might be serious. These terms were ignored on the assumption that the relative velocity of the ion and atom is so slow that electron exchange can frequently occur during a collision. The frequency of exchange is equal to $\Delta E/\hbar$ where ΔE is the energy difference between the symmetrical and antisymmetrical interactions. At large distances ΔE is small and exchange occurs very infrequently. The effect of the relative motion may then be to reduce the magnitude of the exchange interaction at these distances. This would reduce Q_D and hence increase K in (9). It is doubtful however whether the correction arising in this way is large enough. Meyerott (1944) has attempted to obtain some information on this point by calculating the interaction between a lithium ion Li^+ and a helium atom, by quantal methods, and hence obtaining a theoretical value for the mobility of Li^+ in helium to compare with the observations of Hoselitz (1940). The lithium ion has a similar electronic structure to that of helium but the exchange effect is small and better agreement with experiment should be obtained if it is this effect which has been miscalculated with He^+. Only moderate agreement was found, as is clear from reference to Table 2, so that no definite conclusions can be drawn.

Table 2. Comparison of Observed and Theoretical Values for the Mobility of Li^+ in He

Temp. (° K.)		20	78	90	195	291	389	483
Mobility (cm^2/volt. sec.)	Observed	20	21·8	22·2	23·9	25·8	27·8	29·2
	Calculated	16·4	17·4	17·5	18·6	19·4	20·1	20·9

To clarify the position the mobility of He^+ ions with established identity

should be measured. A further possibility would be to measure the coefficient of diffusion of metastable atoms of helium in helium (see Dorrestein and Smit 1938), as in this case also electron and excitation transfer must play an important rôle. The theoretical importance of a satisfactory solution is considerable because these cases of charge and excitation transfer, in which no change in relative kinetic energy occurs, can be regarded as limiting cases of collisions in which charge and excitation transfer can only be effected if the energy of relative motion is changed. A satisfactory theory of these elastic processes must be expected to precede development of such a theory for the more complicated inelastic cases.

§3. COLLISIONS INVOLVING CHANGE OF INTERNAL ENERGY

A considerable volume of evidence is available concerning the ease with which energy exchange may occur between that of relative translation and various degrees of internal freedom of the colliding systems. Nevertheless, although it is possible to give general rules concerning these energy transformations, we are still far from possessing an interpretation so complete as to make possible the prediction of the rates of a particular reaction except in very few cases.

3.1. *Inelastic Collisions between Atoms—Exchange between electronic and translational Energy*

In collisions between atoms, the only internal motion which can exchange energy with that of relative translation is that of the electrons. As we are concerned with impacts in which the relative kinetic energy is of the order kT, inelastic collisions between atoms cannot occur unless one of the atoms is in an excited state.

The occurrence of inelastic (and superelastic) collisions between excited atoms leads to the phenomena of the quenching of resonance fluorescence by the presence of foreign atoms, and of sensitized fluorescence.

Resonance fluorescence is produced when a gas or vapour of atoms A is excited by radiation, from a suitable lamp, which arises from transitions from an excited (nth) state to the ground state of the atoms A in the lamp. Resonance absorption of this radiation raises atoms A in the resonance chamber to the nth state, from which they return to lower states by emission of the resonance radiation. The intensity of this resonance radiation for a given intensity of irradiation from the lamp is often reduced very considerably, or quenched, by the addition of foreign gas to the resonance chamber. This is because the atoms of this gas, in collision with excited atoms A, remove the excitation energy. With suitable experimental precautions, necessary to avoid complications due to imprisonment of resonance radiation and to line broadening by the foreign gas, it is possible to measure the effective cross-section for the quenching due to different atoms and molecules (see Mitchell and Zemansky 1934a, b).

Cross-sections for the quenching of the resonance radiation of mercury (Zemansky 1930, Bates 1930, 1932, Duffendack and Owens 1934, Evans 1934, Olsen 1941, Olsen and Kerr 1947), cadmium (Lifson and Mitchell 1935), thallium (Prileschajewa 1932, 1935), and sodium (Winans 1930, Kisilbasch, Kondratjew and Leipunsky 1930, Terenin and Prileschajewa 1931, Kondratjew and Susken

1935, Norrish and Smith 1940), by different gases have been measured. The data for quenching by molecules are discussed in § 3.2. For all the atomic gases investigated the quenching cross-sections have been found to be very small ($<10^{-17}$ cm^2). Some authors have given actual values for some of the cross-sections but usually it is merely said that they are too small to be observed with the existing experimental limitations. Thus Olsen (1941) finds negligible quenching of mercury radiation by helium and krypton but gives cross-sections 3 and 2×10^{-17} cm^2, for quenching by neon and argon respectively. It seems quite certain, by comparison with the data for quenching by molecules, that atoms are very ineffective in quenching, the chance per gas kinetic collision being much less than 1/10.

Sensitized fluorescence occurs when the irradiation of a mixture of gases A and B, by radiation which excites the atoms A by resonance absorption, leads to emission of radiation by the atoms B. This is due to transfer of excitation from the atoms A to the atoms B in collisions (see Mitchell and Zemansky 1934b). The most important general rule, which has emerged from the experimental study of these phenomena, is that transfer of excitation is most effective when the minimum amount of energy is transferred to or from relative translation, i.e. when the energy difference between the initial and final states of the first atom is equal to that between the final and initial states of the second. In these circumstances the colliding systems are said to be in energy resonance. A most definite illustration of this resonance effect was provided by some experiments of Beutler and Josephy (1929) who investigated the mercury sensitized fluorescence of sodium. They determined the relative population of different excited levels of sodium and found a sharp maximum for the 9s level. The excitation energy of this level is only 0·020 ev. higher than that of the $6\,^3P_1$ level of mercury from which the excitation was transferred. A small subsidiary maximum was also observed for the 7s level whose excitation energy lies closest to that of the metastable $6\,^3P_0$ level of mercury.

Similar resonance effects have been observed in other reactions. Thus Gran and Duffendack (1937) observed the enhancement of the s, p and d levels of the Pb^+ spark lines when neon was added to an electric discharge in lead vapour. The enhancement is due to the transfer of charge and excitation between neon and lead by the reaction

$$Ne^+ + Pb \rightarrow Pb^{+\prime} + Ne. \qquad \ldots\ldots(13)$$

It was found, again, that the maximum enhancement in any series of lines, starting from levels with the same azimuthal quantum number, occurred when the energy resonance was closest.

A second general rule, due to Wigner (1929), is that the total electronic spin is conserved in a transfer reaction. Although this rule is useful as a general guide to the most important processes it is not accurately valid. The most remarkable cases in which it fails occur with helium atoms. In the course of an investigation of the excitation of the helium spectrum by electron impact Lees and Skinner (1932) produced strong evidence to show that the cross-section for the reaction $He(4\,^1P) + He(1\,^1S) \rightarrow He(^1S) + He(4\,^3D) + 0{\cdot}006$ ev., is not small compared with the gas kinetic. Maurer and Wolf (1934, 1940) studied these

reactions in great detail and obtained the effective cross-sections (in terms of the gas kinetic) given in Table 3.

Table 3

Reaction	Resonance discrepancy ev.	Cross-section	Reaction	Resonance discrepancy ev.	Cross-section
$3^1P \rightarrow 3\,^3P$	0·079	1·4	$5^1P \rightarrow 5\,^1D$	0·003	34
$3^1P \rightarrow 3\,^3D$	0·013	8	$5^1P \rightarrow 5\,^3D$	0·003	18
$4^1P \rightarrow 4\,^1D$	0·006	45	$5^1P \rightarrow 5\,^1S$	0·034	<1
$4^1P \rightarrow 4\,^3D$	0·006	10	$5^1P \rightarrow 5\,^3S$	0·071	$<0·6$
$4^1P \rightarrow 4\,^1S$	0·068	<2			
$4^1P \rightarrow 4\,^3S$	0·148	<1			

These results exhibit the resonance behaviour quite clearly but Wigner's rule only holds in a weak form. The cross-section for a process in which the total spin is conserved is larger than for the corresponding process in which it is not, but only by a factor of 2 or 3.

A further source of information about atomic collisions in which electron transitions occur is from the study of the formation and destruction of metastable atoms in an electrically excited gas. These experiments are rather difficult to interpret accurately because a number of factors are involved. An account of experiments in the rare gases have been given by Mitchell and Zemansky (1934a) in *Resonance Radiation and Excited Atoms*. As an example it is found (Meissner and Graffunder 1927, Zemansky 1929) that the cross-section for excitation of metastable neon atoms in either one of the ^3P levels to the $^1P_1$ level which lies 0·15 ev. above the uppermost, $^3P_0$, level, by impact of normal neon atoms at room temperature, is $4·2\times10^{-19}\gamma$ cm^2. γ, the fraction of neon atoms possessing sufficient energy to produce the excitation, is likely to be of order 10^{-2} so that the cross-section is between 0·1 and 0·01 of the gas kinetic. A somewhat smaller but comparable result has been found by de Groot and Penning (1933) from an analysis of some earlier experiments by Dorgelo and Washington (1927). The value they derive is $3·9\times10^{-20}/\gamma$. For argon the cross-section for the excitation $^3P_0 \rightarrow {}^1P_1$ appears to be considerably smaller (Anderson 1931). On the other hand, doubt is thrown on the validity of the method by the impossibly large cross-section found (Ebbinghaus 1930) for the excitation $2\,^3S \rightarrow 2\,^1S_0$ in helium.

Recently Kvifte and Vegard (1947) have reported experimental estimates of the cross-sections for de-activation of oxygen atoms in the metastable ^1S state by impact of neon atoms. In this case the process involves a superelastic collision and the factor γ does not arise. The very low probability of 10^{-7} per gas kinetic collision was found for this process which involves transfer of 2·0 ev. internal energy into energy of relative translation.

Finally, Kruithof and Druyvesteyn (1937) have obtained an estimate of the cross-section for the reaction

$$Ne' + A \rightarrow Ne + A^+ + e, \qquad \ldots\ldots(14)$$

Ne′ being a metastable atom, from a study of the electron ionization coefficient of neon containing a small admixture of argon. This gave the ratio of the cross-section for the above reaction to that for destruction of metastable neon atoms in collision with normal neon atoms. Using de Groot and Penning's value (1933)

for the latter cross-section, the cross-section for the reaction (14) was found to be of the order of the gas kinetic. It is important to notice that the reaction (14) can be regarded as one in which exact resonance may prevail, since the energy excess may be carried by the ejected electron.

3.11. *Theoretical interpretation.* The outstanding feature of the data described above, meagre as they undoubtedly are, is that the cross-section for a collision, in which an electron transition takes place involving a change ΔE of internal energy, is small unless $|\Delta E|$ is very small. In many cases the variation of the cross-section for a fixed relative velocity of impact follows a resonance curve, with a maximum at $\Delta E = 0$.

A qualitative interpretation of this feature may readily be given in terms of the extent to which the collision approximates to adiabatic conditions. Under such conditions the atoms approach so gradually that the state of internal motion is able to adjust itself to the perturbation without a transition finally resulting. We denote the initial state by A, the final state, the energy of which differs from state A by ΔE, by B. When the systems are at a given finite distance apart the state of the system will fluctuate between being mainly like A and mainly like B with a frequency of order $\Delta E/\hbar$. If the number of such fluctuations during the collision is high, the conditions are nearly adiabatic and the chance of finding the systems in state B after the collision is very small. The time of collision is of order a/v, where v is the relative velocity, and a length of the order of atomic dimensions. Hence if

$$\Delta E a/\hbar v \gg 1, \qquad \ldots\ldots(15)$$

the conditions are nearly adiabatic. For a considerable chance of a transition we must have $\Delta E a/\hbar v \simeq 1$ or greater. In the atomic collisions we have been considering $\Delta E a/\hbar v$ is of the order $10\Delta V$ where ΔV is the energy difference in electron volts. We would therefore expect a small cross-section unless ΔV is smaller than 0·1 ev.

The variation of the cross-section with relative velocity of impact will be given by $f(\Delta E a/\hbar v)$ where $f(x)$ is a decreasing function of x. In some circumstances it is likely that the function has the form $e^{-\alpha x}$, in other cases it may be better represented by x^{-n} where $n>0$. Figure 5 illustrates the general form of the variation with velocity.

It is clear, according to the above considerations, that the maximum cross-section occurs when $\Delta E = 0$. The magnitude which can be attained under conditions approaching exact resonance is a matter of some importance.

There is a general impression that very large cross-sections, much greater than the gas kinetic, can arise in collisions involving transfer of excitation with small energy discrepancy ΔE. It seems likely from theoretical considerations, however, that such large cross-sections are the exception rather than the rule. The so-called resonance effect which is a feature of atomic collisions is essentially different in character from the resonance effects which occur in nuclear collisions or in the collisions of slow electrons with atoms. In the latter cases, the wavelength of the relative motion is large compared with nuclear or atomic dimensions respectively, whereas in the atomic case the wavelength is small compared with atomic dimensions and the relative motion is essentially classical. Large cross-sections can only arise in these circumstances if the effective interaction between the atoms has an abnormally long range—in nuclear (or electron) collisions, however, the collision radius may greatly exceed the nuclear (or atomic) radius.

The case of exact resonance, which may be dealt with accurately, provides a good guide to the conditions under which the cross-section for a given electronic transition may become large. As already discussed in § 2.4, in connection with the effect of charge transfer on the mobility of He^+ ions in helium, the cross-section is determined by a suitable weighting of the scattered amplitudes arising from the two interactions between the atoms, one symmetrical, the other antisymmetrical, in the two nuclei. These interactions fall off as r^{-3} if the change of excitation involves an optically allowed transition in both atoms, with no electron exchange. This long range interaction can give rise to a very large cross-section if its strength is great enough. If, however, one or both of the transitions are not optically allowed, the interactions fall off more rapidly with nuclear separation and large cross-sections do not occur. In particular, if charge exchange takes place, the interaction falls off exponentially and a cross-section much larger than gas kinetic cannot arise (except when the colliding systems are oppositely charged ions (see equation (16)).

These general conclusions remain valid when the resonance is not exact. Essentially this has the effect of reducing the effective long range interaction by interference between the waves of different length corresponding to the initial and final relative motion. The most detailed theory of such collisions has been given by Stueckelberg (1932) (see also Landau 1932, London 1932, Zener 1932, Mott and Massey 1949), but it has not been applied to specific examples except in very few cases. One rather exceptional reaction to which it has been applied is the mutual neutralization reaction

$$A^+ + B^- \rightarrow A' + B''. \qquad \ldots\ldots (16)$$

The very long range Coulomb force between the colliding ions can, under certain circumstances, lead to a cross-section for the reaction which is much greater than the gas kinetic. Bates and Massey (1943) applied Stueckelberg's theory to estimate the possible cross-section for mutual neutralization, at ordinary temperatures, of the positive and negative ions of atomic oxygen, a process which may be of importance in the earth's ionosphere. They found that, in these cases, the maximum cross-section does not occur at exact resonance but when the resonance discrepancy is one or two electron volts. It is possible that a cross-section as high as 10^{-12} cm^2 might arise for some pair of excited states in (16) but this would require very special conditions. The actual value is likely to be somewhat smaller but still rather larger than the gas kinetic.

Although a general theoretical interpretation of the semi-adiabatic reactions between excited atoms or ions can be given we are still far from possessing a means of estimating even roughly the magnitudes of the cross-sections for any specific conditions. More experimental data, particularly on the variation of the cross-sections with velocity, must be secured to assist in the difficult task of providing at least a semi-quantitative or semi-empirical theory which is useful in the frequently occurring circumstances in which the rate of a reaction, unobtainable by experiment, is required.

3.2. *Inelastic Collisions between Atoms and Molecules—Exchange between electronic, vibrational and translational Energy*

In considering collisions involving electron transitions in which one or both of the colliding systems are molecules, account must be taken of the possibility that vibrational motion within the molecule may give or receive energy. One

immediate question of interest is whether or not vibrational energy can be considered on the same footing as electronic energy in assessing the resonance energy discrepancy. Experimental evidence on this matter can be obtained by studying the quenching of resonance fluorescence by molecular gases. Care must be taken to distinguish quenching by chemical reactions from direct quenching by transfer of excitation.

There is no doubt from the observed data that electronic energy may, in many cases, be transformed quite readily into vibrational energy. Thus the cross-section for the quenching of sodium resonance radiation by hydrogen is $7{\cdot}5\times10^{-16}$ cm² (Norrish and Smith 1940). This is much larger than the quenching by rare gas atoms and the only explanation of this is the possibility, with hydrogen, that the electronic energy be transferred into energy of molecular vibration. Again, in the quenching of mercury resonance radiation by hydrogen, the effective process has been found to be

$$Hg(^3P_1)+H_2\rightarrow Hg(^1S_0)+2H, \qquad \ldots\ldots(17)$$

with the corresponding cross-section about 6×10^{-16} cm² (Olsen and Kerr 1947, Zemansky 1930). In this case the vibrational excitation is great enough to lead to dissociation. Many chemical reactions are sensitized by excited mercury atoms (see, for example, Noyes and Leighton 1941) due to this convertibility of electronic into vibrational energy. Dissociation of hydrogen by excited xenon atoms has also been observed (Calvert 1932). On the other hand, the quenching of cadmium resonance radiation by hydrogen takes place mainly through the reaction

$$Cd(^3P_1)+H_2\rightarrow CdH+H, \qquad \ldots\ldots(18)$$

the cross-section being about 6×10^{-16} cm² (Lifson and Mitchell 1935). In this case the transfer of electronic to vibrational energy must take place with greater difficulty.

It is not clear to what extent the size of the cross-section depends on the difference between the electronic excitation energy and that of the vibrational excitation to which it is wholly or partly transferred. It was first thought (Zemansky 1930) that the observed cross-sections for quenching of mercury resonance radiation by molecular gases established quite definitely that the most effective molecules are those which possess an excited vibrational state with excess energy most nearly equal to the energy difference (0·218 ev.) between the 6^3P_1 and the metastable 6^3P_0 states of the mercury atom. Subsequent investigation has shown that many of the more efficient quenchers, such as hydrocarbons, act through chemical changes. The evidence based on the remaining molecules, which are known to quench by producing a transition to the metastable state, is much less definite and the existence of a resonance effect is still doubtful.

The first attempt at a detailed interpretation of data on quenching by molecules was made by Magee and Ri (1941). They applied the transition state method (see Glasstone, Laidler and Eyring 1941) to discuss the quenching of sodium resonance radiation by hydrogen. The method consists in tracing the detailed course of the reaction in terms of the potential energy surfaces for different electronic states. An electronic transition may occur at an intersection of two such surfaces. The location of these intersections, together with the nature of the electronic states concerned and the general shape of the potential energy surface, determines the rate of the quenching reaction for given initial conditions. The

procedure is purely classical except in so far as the probability of a transition at the crossing point is concerned. A more extensive study on these lines has been made by Laidler (1942, 1943). Despite the complexity of the problem he was able to provide a satisfactory description of the quenching of sodium by hydrogen and also, to some extent, by the halogens, obtaining in particular quite good estimates of the effective cross-sections. There is nothing in his description which suggests that a resonance effect would be expected.

3.3. *Inelastic Collisions between Atoms and Molecules—Exchange between vibrational and translational Energy*

The rate of exchange of energy between molecular vibration and translation is of importance in a number of physical phenomena. Thus it determines the critical pressure in a unimolecular reaction. Again, the persistence of vibration in collisions is responsible for the dispersion of sound and may even influence the high speed flow of gases.

An argument (Zener 1931, Landau and Teller 1936) similar to that presented in §3.11 would lead us to expect that the chance of exchange of energy between vibration and translation would be small when $\nu a/v$ is large, ν being the vibration frequency, v the velocity of relative motion and a a length of the order of the gas kinetic radius.

If d is the amplitude of vibration, which will be less than a, $\nu^2 d^2$ is of order $h\nu/m$ where m is the reduced mass of the vibrational motion. The condition for persistent vibration may therefore be written in the form

$$h\nu a^2/kTd^2 \gg 1, \qquad \ldots\ldots (19)$$

since mv^2 will be of order kT, the temperature energy of the gas concerned.

The ratio a/d will be quite large for the lowest vibrational states of most molecules, so that the condition (19) may be satisfied even if the vibrational quantum $h\nu$ is comparable with the mean temperature kinetic energy of the molecules, so the vibrational energy makes an appreciable contribution to the specific heat. On the other hand, the vibrational level separation in a molecule which is excited, say to half the dissociation energy, is usually very much smaller than kT at ordinary temperatures. The ratio a/d is also considerably greater so that the condition (19) may well be violated and easy exchange between vibration and translation expected. The experimental evidence available supports these conclusions.

Persistent vibration leads to a modification of the vibrational contribution to the specific heat in a gas in which rapid temperature changes are occurring. This will be unimportant if the vibrational specific heat is in any case very small, as for hydrogen in which the energy difference between the ground and first excited vibrational state is as high as 0·54 ev. For other molecules, it is effective in producing a dependence on the frequency of the velocity of a high-frequency sound wave. This leads to absorption as well as dispersion because the waves of pressure and density are no longer exactly in phase.

If only one vibrational transition is concerned, the velocity of sound V and absorption coefficient per wavelength α, are given by (Kneser 1931, see also Hiedemann 1935. Richards 1939)

$$\left.\begin{aligned} V^2 &= (V_0^2\omega_i^2 + V_\infty^2\omega^2)/(\omega_i^2+\omega^2), \\ \alpha &= 2\pi(V_\infty^2 - V_0^2)\omega_i\omega/(V_0^2\omega_i^2 + V_\infty^2\omega^2). \end{aligned}\right\} \qquad \ldots\ldots (20)$$

V_0 and V_∞ are the velocities of sound of vanishing and infinitely high frequency respectively. In calculating V_0 the specific heat c_0 includes the full vibrational contribution whereas in V_∞ the appropriate specific heat c_∞ contains no vibrational contribution. $\omega/2\pi$ is the frequency of the sound and $\omega_i = c_0/c_\infty\beta$. $1/\beta$ is the chance per second that a vibrationally excited molecule will be de-activated, so that β is a relaxation time.

According to this theory, β should be proportional to the pressure and might be expected, from the condition (19), to decrease with increasing temperature. If a small amount of foreign gas is added, which is more effective in producing vibrational de-activation than the main gas, the relaxation time is reduced and the dispersion shifted to higher frequencies. A small admixture of foreign gas can never increase the relaxation time and move the dispersive region to lower frequencies. This, and the linearity with pressure, can be used as a test of the applicability of the theory.

A great number of measurements have been made of the dispersion and absorption of sound (Richards 1939, Eucken and Küchler 1938) in pure and mixed gases. The results show that the chance of vibrational de-activation in a collision may be very small indeed. Thus, for pure carbon dioxide (Wallman 1935, Eucken and Becker 1933), it is found to be about 2×10^{-5} per collision at 18° C. The process is also highly specific. The chances of de-activation of excited CO_2 per collision with CO_2, CH_4, He, H_2 and H_2O are as 1 : 18 : 35 : 200 : 400 : 2,000.

As expected, the experimental results show that the chance of de-activation increases quite rapidly with temperature. Thus Eucken and Becker (1933, 1934) found that there was a seven-fold increase in the chance, between temperatures of $-32°$ C. and 145° C. in chlorine, and a four-fold increase between the same two temperatures in carbon dioxide.

According to theoretical prediction (Jackson and Mott 1932, Zener 1932, Landau and Teller 1936) the chance of de-activation per collision is, in the notation of (19), closely proportional to $\exp(-2\pi a\nu/v)$. The fraction of impacts in which the relative velocity is v will be proportional to the Maxwell–Boltzmann factor $\exp(-\frac{1}{2}Mv^2/\boldsymbol{k}T)$ where M is the reduced mass in the collision. Combining these two factors, it is found that the largest contribution to the de-activation process will be due to collisions in which the relative velocity is given by

$$v^3 = 2\pi a\nu\boldsymbol{k}T/M.$$

The fraction of such collisions will be proportional to $\exp\{-(27\pi^2a^2\nu^2M/2\boldsymbol{k}T)^{\frac{1}{3}}\}$ so that the relaxation time should vary with temperature as $\exp(\alpha T^{-\frac{1}{3}})$ where α is a negative constant. This is in general agreement with the observations (see Kantrowitz 1945).

No detailed theoretical interpretation of the wide range of experimental data now available has yet been made, although Franck and Eucken (1933) have attempted to explain some of the specific features by supposing that if the colliding molecule can react chemically with the vibrationally excited molecule it is more effective in de-activation.

The failure of the vibrational degrees of freedom to follow fast temperature changes gives rise to observable effects in the rapid flow of gases. In such flow, temperature changes will occur at a rate depending on the dimensions of the obstacles encountered and on the velocity of the flow. If the rates of change

are comparable with the vibrational times, departure from the equilibrium partition of energy within the gas will occur. This will lead to irreversible flow of heat from those degrees of freedom which possess more than their equilibrium share to those with less, thereby increasing the entropy. It is possible to determine relaxation times by measuring the entropy increase under suitable flow conditions. This has been done by Kantrowitz (1945) for carbon dioxide. His results agree well with those given by the sonic measurements. The effect may arise in high speed gas turbines and must be taken into account in choosing gases for high speed flow (such as wind-tunnel) experiments.

The occasional disagreement of specific heats measured by explosion methods with those obtained in other ways has also been traced to vibrational heat lag (Wohl and Magat 1932).

A direct spectroscopic verification of the persistence of vibration in the first excited vibrational state has been carried out by Dwyer (1939). The absorption bands of iodine in the visible arise, under normal conditions, from transitions from the ground vibrational state. If the iodine is excited by an electric discharge, an appreciable population of iodine molecules in the first excited vibrational state is produced by downward transitions from higher electronic states. The absorption bands given by this excited iodine show a greater intensity due to transitions from the first excited vibrational states than from the ground state. If the discharge is cut off suddenly, it is found that the enhancement of these bands persists for a time of order 1/30 sec. during which the average molecule suffers 7,000 collisions.

Experimental evidence supporting the theoretical expectation that persistence of vibration is not apparent when highly excited vibrational quanta, much smaller than kT, are involved, has been provided from studies of the sensitized band fluorescence of molecules. Rossler (1935) studied the effect of rare gases on the band fluorescence of I_2 excited by the green mercury line 5461 A. Absorption of this line excites the molecule to an electronic state with vibrational quantum number as high as 26. The foreign gas added not only produces quenching but also leads to the emission of new bands. These arise from vibrational levels excited by collisions with the foreign atoms. A quantitative study showed that the cross-sections for vibrational excitation are certainly of the order of the gas kinetic, in sharp contrast to the results obtained from the dispersion of sound. Similar results were obtained by Durand (1940) in an investigation of the fluorescence of sulphur excited, by absorption of a magnesium spark line, to an electronic state in which the excess vibrational energy was $\frac{1}{4}$ of the dissociation energy.

A unimolecular reaction is really a bimolecular one in which the time taken for the vibrationally excited molecule to break up is long compared to the time between collisions in which vibrational transitions occur. As the pressure is reduced the true bimolecular character of the reaction will become manifest. The critical pressure at which this becomes apparent will depend on the chance per collision of a vibrational transition. For a particular decomposing molecule, the critical pressure is found (Oldenburg and Frost 1937) to depend very markedly on the nature of the admixed gas, showing again that the chance of vibrational activation on collision is very specific. Absolute values of the chance per collision cannot be obtained with any certainty from the critical pressure. The degree of vibrational excitation in the activated molecule will vary greatly according

to the mode of internal motion excited, so it is to be expected that the average chance should be roughly intermediate between the very low values found in the ultrasonic experiments and the values of order unity given from the band fluorescence experiments.

3.4. *Inelastic Collisions between Atoms and Molecules—Exchange between rotational and translational Energy*

In all gases, except hydrogen at very low temperatures, the conditions for exchange of rotational and translational energy are the reverse of adiabatic, the time of collision being short compared with the time of rotation. If a is the gas kinetic radius, the angular velocity ω of the rotation is of order $\boldsymbol{h}/2\pi Ma^2$, M being the reduced mass of the molecule, so that the time of rotation is of order $4\pi^2Ma^2/\boldsymbol{h}$. For a nearly adiabatic collision we must have then

$$4\pi^2Ma^2/\boldsymbol{h} \ll a/v, \quad \text{or} \quad 4\pi^2Mav/\boldsymbol{h} \ll 1,$$

whereas, with the exception noted, the opposite inequality holds. We would therefore expect that rotational excitation or de-activation should occur readily on impact.

This viewpoint is supported by evidence from experiments such as those of Rossler (1935) and of Durand (1940) discussed above. The effect of the foreign gas provides a broadening of the fluorescent lines which can only be ascribed to the excitation of many rotational states. A rough estimate of the cross-section for these excitations shows that it is of the same order as the gas kinetic.

Further evidence has been obtained by studying the band emission from molecules possessing abnormal rotation when they are first formed by some dissociation process. Thus Oldenberg (1931, 1934) has studied the abnormal rotation of OH radicals produced on dissociation of water molecules by electrons. He finds that it disappears quite quickly when helium or argon is added. On the other hand, in the most recent observations of Oldenberg and Rieke (1938), a semi-quantitative estimate has indicated that the rotation can persist through some 500 or more collisions.

van Itterbeek and Mariens (1940) have reported measurements of sound absorption in hydrogen and deuterium which give a relaxation time for the rotation of order 10^{-8} sec. This requires a cross-section for rotational de-activation of about 10^{-17} cm^2 which is of the order expected (see Roy and Rose 1935), remembering that at the temperature of the observations (0°–90° c.) the collisions will be far from adiabatic. Instead the cross-section may be reduced because the time of impact is so short compared with that of a complete rotation.

Interesting phenomena arising from the redistribution of energy between vibration and rotation on impact have been observed by various authors. Thus abnormally high rotation has been observed in the band fluorescence of HgH excited by the 2537 mercury line when molecular nitrogen is present. The HgH molecules are formed with excess vibrational energy. In collisions with nitrogen molecules, energy exchange with the light hydrogen atom takes place much more readily than with the heavy HgH molecule as a whole, so that change of vibration into rotation can occur, although there is little change of relative translational energy in the average impact. In this, as in many other similar examples, a classical description is possible.

REFERENCES

Amdur, I., 1947, *J. Chem. Phys.*, **5**, 482.
Amdur, I., and Pearlman, H., 1941, *J. Chem. Phys.*, **9**, 503.
Anderson, J. M., 1931, *Can. J. Res.*, **4**, 312.
Bates, D. R., and Massey, H. S. W., 1943, *Phil. Trans. Roy. Soc.* A, **239**, 269.
Bates, J. R., 1930, *J. Amer. Chem. Soc.*, **52**, 3825; 1932, *Ibid.*, **54**, 569.
Beth, E., and Uhlenbeck, G. E., 1937, *Physica*, **4**, 915.
Beutler, H., and Josephy, B., 1929, *Z. Phys.*, **53**, 755.
de Boer, J. H., and Michels, A., 1939, *Physica*, **6**, 409.
Broadway, L. F., 1933, *Proc. Roy. Soc.* A, **141**, 634.
Buckingham, R. A., 1938, *Proc. Roy. Soc.* A, **168**, 264.
Buckingham, R. A., Hamilton, J., and Massey, H. S. W., 1941, *Proc. Roy. Soc.* A, **179**, 103.
Calvert, H. R., 1932, *Z. Phys.*, **78**, 479.
Chapman, S., 1916, *Phil. Trans. Roy. Soc.* A, **216**, 279; 1917, *Ibid.*, **217**, 115.
Cuthbertson, C., and Cuthbertson, M., 1911, *Proc. Roy. Soc.* A, **84**, 13.
Dorgelo, H. B., and Washington, H. S., 1927, *Proc. K. Acad. Amst.*, **30**, 33.
Dorrestein, S., and Smit, J. H., 1938, *Proc. K. Acad. Amst.*, **41**, 725.
Duffendack, O. S., and Owens, J. S., 1934, *Phys. Rev.*, **46**, 417.
Durand, E., 1940, *J. Chem. Phys.*, **8**, 46.
Dwyer, R. J., 1939, *J. Chem. Phys.*, **7**, 40.
Ebbinghaus, E., 1930, *Ann. Phys. Lpz.*, **7**, 267.
Enskog, D., 1917, *Inaug. Diss., Upsala.*
Estermann, I., Foner, S. N., and Stern, O., 1947, *Phys. Rev.*, **71**, 250.
Eucken, A., and Becker, R., 1933, *Z. phys. Chem.* B, **20**, 467; 1934, *Ibid.*, **27**, 219.
Eucken, A., and Küchler, L., 1938, *Z. tech. Phys.*, **19**, 12, 517.
Evans, M. G., 1934, *J. Chem. Phys.*, **2**, 445.
Franck, J., and Eucken, A., 1933, *Z. phys. Chem.* B, **20**, 467.
Glasstone, S., Laidler, K. J., and Eyring, H., 1941, *The Theory of Rate Processes* (London and New York: McGraw Hill), Chaps. IV, VI.
Gran, W. H., and Duffendack, O. S., 1937, *Phys. Rev.*, **51**, 804.
Gropper, L., 1937, *Phys. Rev.*, **51**, 1108.
de Groot, W., and Penning, F. M., 1933, *Handbuch der Phys.*, XXII, 2nd edition, p. 167.
Haber, H., 1939, *Phys. Z.*, **40**, 541.
Hassé, H. R., and Cook, W. R., 1931, *Phil. Mag.*, **12**, 554.
Hellmann, H., 1935, *Acta Phys. Chem. U.S.S.R.*, **2**, 273.
Hiedemann, E., 1935, *Ergebn. exakt. Naturw.*, **40**, 201.
Hirschfelder, J. O., Ewell, R. B., and Roebuck, J. R., 1938, *J. Chem. Phys.*, **6**, 205.
Holborn, L., and Otto, J., 1926, *Z. Phys.*, **38**, 359.
Hoselitz, K., 1940, *Proc. Roy. Soc.* A, **177**, 200.
van Itterbeek, A., and Keesom, W. H., 1938, *Physica*, **5**, 257.
van Itterbeek, A., and Mariens, P., 1940, *Physica*, **7**, 938.
Jackson, J. M., and Mott, N. F., 1932, *Proc. Roy. Soc.* A, **137**, 703.
Kantrowitz, A., 1945, *J. Chem. Phys.*, **14**, 150.
Keesom, W. H., and Kraak, H. H., 1935, *Physica*, **2**, 37.
Keesom, W. H., and Walstra, W. K., 1939, *Physica*, **2**, 257.
Kisilbasch, B., Kondratjew, V., and Leipunsky, A., 1930, *Phys. Z. Sowjet*, **2**, 201.
Kneser, H. O., 1931, *Ann. Phys. Lpz.*, **11**, 761.
Kondratjew, V., and Susken, I., 1935, *Phys. Z. Sowjet*, **8**, 6, 644.
Kruithof, A. A., and Druyvesteyn, M. J., 1937, *Physica*, **4**, 450.
Kvifte, H., and Vegard, L., 1947, *Geofys. Publ.* XVII, 1.
Laidler, K. M., 1942, *J. Chem. Phys.*, **10**, 34; 1943, *Ibid.*, **10**, 43.
Landau, L., 1932, *Phys. Z. Sowjet*, **2**, 46.
Landau, L., and Teller, E., 1936, *Phys. Z. Sowjet*, **10**, 34.
Lees, W. S., and Skinner, H. W., 1932, *Proc. Roy. Soc.* A, **137**, 186.
Lennard-Jones, J. E., 1931, *Proc. Roy. Soc.* A, **43**, 461.
Lipson, H. C., and Mitchell, A. C. G., 1935, *Phys. Rev.*, **48**, 625.
London, F., 1932, *Z. Phys.*, **74**, 143.

MAGEE, J. L., and RI, T., 1941, *J. Chem. Phys.*, **9**, 638.
MARGENAU, H., 1939, *Phys. Rev.*, **56**, 1000.
MASSEY, H. S. W., and BUCKINGHAM, R. A., 1936, *Nature, Lond.*, **138**, 77; 1938, *Proc. Roy. Soc.* A, **168**, 378.
MASSEY, H. S. W., and MOHR, C. B. O., 1933, *Proc. Roy. Soc.* A, **141**, 434; 1934, *Ibid.*, **144**, 188.
MASSEY, H. S. W., and SMITH, R. A., 1933, *Proc. Roy. Soc.* A, **142**, 142.
MAURER, W., and WOLF, R., 1934, *Z. Phys.*, **92**, 100; 1940, *Ibid.*, **115**, 410.
MEISSNER, K. W., and GRAFFUNDER, W., 1927, *Ann. Phys. Lpz.*, **84**, 1009.
MEYEROTT, R., 1944, *Phys. Rev.*, **66**, 242.
MITCHELL, A. C. G., and ZEMANSKY, M. W., 1934a, *Resonance Radiation and Excited Atoms* (Cambridge : University Press), Chapter IV; 1934b, *Ibid.*, Chapter II.
MOTT, N. F., 1930, *Proc. Roy. Soc.* A, **127**, 658.
MOTT, N. F., and MASSEY, H. S. W., 1949a, *The Theory of Atomic Collisions*, Oxford, 2nd edition, Chapter VII, § 2; 1949b, *Ibid.*, Chapter XII, § 3.3.
NIJHOFF, G. P., ILIIN, B., and KEESOM, W. H., 1928, *Comm. Phys. Lab. Univ. Leiden*, No. 188c.
NORRISH, R. W., and SMITH, W. M., 1940, *Proc. Roy. Soc.* A, **176**, 295.
NOYES, W. A., and LEIGHTON, P. A., 1941, *Photochemistry of Gases* (New York : Reinhold), Ch. V.
OLDENBERG, O., 1931, *Phys. Rev.*, **37**, 1550; 1934, *Ibid.*, **46**, 210.
OLDENBERG, O., and FROST, A. A., 1937, *Chem. Rev.*, **20**, 99.
OLDENBERG, O., and RIEKE, F. F., 1938, *J. Chem. Phys.*, **6**, 439.
OLSEN, L. O., 1941, *Phys. Rev.*, **60**, 739.
OLSEN, L. O., and KERR, G. P., 1947, *Phys. Rev.*, **72**, 115.
PRILESCHAJEWA, N., 1932, *Phys. Z. Sowjet*, **2**, 351, 367; 1935, *Acta Phys. Chem.*, **2**, 51, 647.
RICHARDS, W. T., 1939, *Rev. Mod. Phys.*, **11**, 36.
RIEKE, F. F., 1936, *J. Chem. Phys.*, **4**, 513; 1937, *Ibid.*, **5**, 831.
ROSIN, S., and RABI, I. I., 1935, *Phys. Rev.*, **48**, 373.
RÖSSLER, F., 1935, *Z. Phys.*, **96**, 251.
ROY, A. S., and ROSE, M. E., 1935, *Proc. Roy. Soc.* A, **149**, 511.
SCHEFFERS, H., and STARK, J., 1934, *Phys. Z.*, **35**, 625.
SLATER, J. C., 1928, *Phys. Rev.*, **32**, 349.
STUECKELBERG, E. C. G., 1932, *Helv. Phys. Acta*, **5**, 370.
TERENIN, A., and PRILESCHAJEWA, N., 1931, *Z. phys. Chem.* B, **13**, 72.
TYNDALL, A. M., 1938, *Mobility of Positive Ions in Gases* (Cambridge : University Press).
TYNDALL, A. M., and POWELL, C. F., 1930, *Proc. Roy. Soc.* A, **129**, 162; 1931, *Ibid.*, **134**, 125.
UBBINK, J. B., and DE HAAS, W. J., 1943, *Physica*, **10**, 645.
WALLMAN, M-H., 1935, *Ann. Phys., Lpz.*, **21**, 671.
WERGELAND, H., 1945, *Skrif. Norske Videns. Akad. Oslo, Mat.-Nat. Klasse*, No. 9.
WIGNER, E., 1929, *Gött. Nachr.*, 375.
WINANS, J. G., 1930, *Z. Phys.*, **60**, 631.
WOHL, K., and MAGAT, M., 1932, *Z. phys. Chem.*, **19**, 117.
ZEMANSKY, M. W., 1929, *Phys. Rev.*, **34**, 213; 1930, *Ibid.*, **36**, 919.
ZENER, C., 1931, *Phys. Rev.*, **37**, 556; 1932, *Proc. Roy. Soc.* A, **137**, 696; 1933, *Proc. Camb. Phil. Soc.*, **29**, 136.

LOW TEMPERATURE PHYSICS

By K. MENDELSSOHN
The Clarendon Laboratory, Oxford

CONTENTS

	PAGE
§ 1. Introduction	270
§ 2. Advances in technique	270
§ 3. Superconductivity	273
§ 4. Liquid helium	280
§ 5. Helium 3	286

§ 1. INTRODUCTION

MODERN low temperature research covers a wide field of physical phenomena. Since before the war the number of laboratories engaged in this work has increased by a factor of two or three, resulting in a corresponding increase in publications. The general interest in low temperature phenomena is also reflected in the ever growing amount of theoretical work. It is, under these circumstances, quite impossible to deal with all aspects of the subject in a single report. The present Report is therefore confined to recent advances in technique and to the two physical phenomena peculiar to low temperatures: superconductivity and superfluidity. These three headings cover the bulk of the published work, but they leave undiscussed, apart from many isolated investigations on conductivity of heat and electricity, specific heats, adsorption, etc., such items as the magnetic cooling method and temperature scale. These subjects, on which no previous reports have been published, cannot be adequately dealt with under sub-headings of the present survey*.

§ 2. ADVANCES IN TECHNIQUE

The most outstanding contribution to the technique of low temperatures which has been made in the last few years is the development of a helium liquefier by Collins (1947). The importance of his refrigeration engine is not to be found in a novel principle or even a new method, but in the fact that for the first time the construction of such a liquefier has been tackled as an engineering problem. Cooling is produced by an expansion engine like that introduced by Kapitza (1934), the last stage of actual liquefaction, also similar to Kapitza's method, being effected by Joule–Thomson expansion. However, there are certain features in Collins' liquefier which remove it from a laboratory apparatus to a piece of industrial equipment. The most important of these are close-fitting pistons which, like the cylinders, are made of nitrided alloy steel of extreme hardness, and flexible piston rods. The rods are not pushed, but operate under tension, thereby allowing perfect self-alignment of the piston inside the cylinder. In this way friction is reduced to a minimum. The work on cooling engines was started by the author as early as 1935 and had in 1939 yielded a diaphragm engine of interesting construction which would maintain a temperature of 10° K. With the outbreak of war the efforts were turned towards the development of mobile

* It is hoped to cover these subjects in subsequent volumes.—Ed.

oxygen expanders which were all piston-cylinder-crankshaft assemblies and which led ultimately to the construction of the helium engine. The liquefier is now being produced by the Arthur D. Little Company and is commercially available. Either experiments can be carried out inside the liquefier, or liquid helium condensed inside the apparatus from a cylinder can be siphoned out into an external cryostat. Starting from room temperature the liquefier will produce helium after four hours, a time which can be reduced to two hours by pre-cooling the helium cycle with liquid air.

In 1932 Simon introduced an easy method of liquefying helium by expanding adiabatically the gas from a pressure cylinder which had been cooled with liquid or solid hydrogen. The expansion is carried out by simply allowing the gas to escape into a gas-holder. There has been some uncertainty of data concerning the yield of the method (van Itterbeek 1939, van Itterbeek, van Paemel and Mariens 1943), which has now been resolved by a systematic investigation by Pickard and Simon (1948) of the yields of liquid under different starting pressures and temperatures. Owing to the specially favourable conditions (low heat capacity of metal container, low density of liquid helium) the yields, expressed as a percentage of cylinder volume filled with liquid, are very high. Using a pressure of 150 atmospheres and a starting temperature of 11° K., 85% yield can be obtained, while with the same pressure at 10° K. the yield will even slightly exceed 100%. This result is, of course, not paradoxical but is due to the low density of the liquid at 4° K. in comparison with that of the compressed gas at 10° K. Comparing the yield of the simple expansion with the calculated yield of a cylinder and piston shows that with this simple method about 60% of the "ideal" yield can be obtained. The large heat capacity of the helium gas in

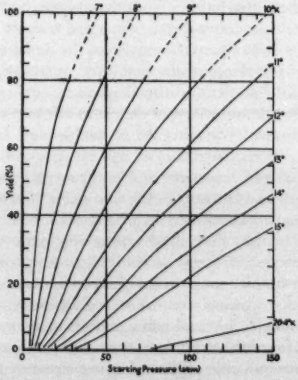


Figure 1. Yield of liquid helium by the expansion method at different starting pressures and temperatures (percentage of volume of expansion cylinder filled with liquid). (*Pickard and Simon.*)

the pressure cylinder renders the expansion method a useful means for obtaining fairly constant temperatures between 14° K. and 4° K., a temperature region which the normal liquid helium cryostat does not cover. Using expansion in this way

in stages MacDonald and Mendelssohn (1948a) have measured electrical resistivities in the intervening temperature region. For similar experiments de Haas and de Boer (1934) used a desorption cryostat (Simon 1926), also with good success.

These examples show that in spite of the possibility of obtaining liquid helium with expansion engines without the use of liquid hydrogen, the latter still plays an important part as a cooling agent. Many arguments can be brought for or against the liquid hydrogen stage in the modern low temperature laboratory. However, there is little doubt that the inclusion of the hydrogen stage will allow greater flexibility of experimental equipment and that the availability of liquid hydrogen will allow experiments which otherwise would hardly be feasible. Some of these factors are discussed by Jones, Larsen and Simon (1948), who describe a hydrogen liquefier producing about 13 litres per hour. The principle used is the conventional Linde–Hampson method, but many novel features are introduced. The chief aim guiding the construction was to produce a liquefier which can be operated regularly and reliably by semi-skilled personnel (research workers). The storage of quantities of liquid hydrogen which have to be used in an economical manner presents a problem which is peculiar to hydrogen and which was not recognized for a long time. Hydrogen gas at ordinary temperatures is composed of a 75/25 per cent mixture of ortho- and para-hydrogen. At the boiling point the equilibrium mixture is 100% para-hydrogen. The conversion rate in the gas is, unfortunately, slow, so that the liquid at the time of its production has the composition of the gas. The conversion is, however, rapid in the liquid state, and the heat of transformation results in an additional evaporation (Larsen, Simon and Swenson 1948). This evaporation is considerable, amounting to 20% of the total stock of liquid in the first 24 hours. This presents a serious problem in cases when quantities of hydrogen have to be stored and used in experiments on several successive days. A partial remedy is to pass the hydrogen in the liquefier over a catalyst (charcoal) at the lowest feasible temperature. This was the reason for the introduction of a charcoal "stage" in the above mentioned liquefier; it has the additional advantage of removing the last traces of gas impurities. By this reduction of the ortho-content of the liquid to 40–45% the evaporation of liquid hydrogen due to conversion can be much reduced: instead of losing 20% in one day this loss will be now spread over five days.

A number of hydrogen liquefiers and equipment used with them has been described by Blanchard and Bittner (1942), DeSorbo, Milton and Andrews (1946), and Fairbank (1946) and a combined hydrogen–helium liquefier by Schallamach (1943). Ranque (1933) has described a heat separator which has produced a certain amount of interest in view of its refrigerating possibilities. The construction is extremely simple, consisting mainly of a tube into which a compressed gas is fed in the middle. The gas stream enters the tube tangentially and is thereby forced into a vortex flow. This whirling motion produces a compression near the wall and an expansion in the centre of the tube. By introducing a diaphragm with a central hole into one side of the tube, the streams of hot and cold gas can be separated. It is clear that such an arrangement is bound to be very inefficient as a refrigerator; its value lies in its simplicity. Ranque, using a pressure of six atmospheres at 20° C., obtained a cooling to −12° C.; Hilsch (1947) obtained a cooling of 68° C. in this way. Johnson (1947), who obtained coolings of 30° C., estimates the efficiency of the device as 15%. A quantitative treatment of the process has been carried out by Burkhardt (1948).

Because of the lower heat capacities at low temperatures the question of heat transfer in calorimeters and cryostats is of particular importance. A systematic study of the transfer by radiation has been carried out by Blackman, Egerton and Truter (1948). It was found that more heat was transmitted from the walls of an enclosure at ordinary temperatures to an inner vessel at 90° K. than can be accounted for by reflectivity of this vessel. The reflectivity had been calculated from the electric conductivity. It was found that a freshly reduced surface of copper had better qualities than a highly polished one for diminishing the inflow of heat. Corrections for incomplete thermal insulation of calorimeters have been worked out by Aston and Szasz (1947).

As mentioned before, a discussion of the magnetic cooling method lies beyond the scope of this Report. For a survey of the general technique developed by Simon and his co-workers, the reader is referred to a summary by Hull (1947).

§ 3. SUPERCONDUCTIVITY

A good deal of experimental as well as of theoretical work on superconductivity has been published since the last Report in this series (Mendelssohn 1946, subsequently referred to as I) appeared. Considerable interest was aroused when Ogg (1946a) reported the observation of superconductivity in solid sodium–ammonia solutions at liquid air temperatures. The work had been carried out on the basis of a theory by the same author postulating that rapidly cooled solutions should contain electrons in the configuration of "trapped pairs" which would obey Bose–Einstein statistics. The Bose–Einstein condensation phenomenon, which because of the small mass of the electrons should in this case occur at fairly high temperatures, was considered to be responsible for the appearance of superconductivity. The actual experiments consisted in determinations of the electrical conductivity when a large decrease in the resistance on cooling was observed, and in observations of persistent currents in ring-shaped specimens. The small residual resistance in the former was ascribed to faulty contact with the platinum electrodes, while as regards the observation of persistent currents "numerous control experiments obviated any other explanation". The phenomenon was investigated by workers in New York (Boorse, Cook, Pontius and Zemansky 1946) and in Oxford (Daunt, Désirant, Mendelssohn and Birch 1946), both repeating the persistent current experiments. The latter also carried out conductivity measurements, using a current potential method. All these tests gave negative results, the conductivity experiments showing also that the finite resistance was not due to faulty contacts but was a feature of the solution. In a second paper Ogg (1946b) stated that of his original 200 experiments on persistent currents only seven had been successful, and in these the currents had disappeared in one or two minutes. The Oxford workers felt that such a rare and transient phenomenon might have escaped their notice and, particularly since meanwhile confirmatory observations had been reported by Hodgins (1946), it was decided to investigate the effect more thoroughly. The new experiments (MacDonald, Mendelssohn and Birch 1947) covered resistivity, persistent currents and susceptibility. Again all results were negative, but it was observed that small magnetic moments such as observed by Hodgins are easily simulated by spurious magnetization in the laboratory. The shape of the resistance curve against temperature can be explained by the appearance of a highly conducting eutectic which has been further investigated (Birch and MacDonald 1947). Similar

conclusions on the basis of further negative results were reached by Giulotto and Gigli (1947). Two subsequent investigations (Gibney and Pearson 1947, Tumanov, Shalnikov and Sharvin 1947) have also failed to confirm the original observation. While it is, of course, impossible to prove that an effect does not exist, the available evidence strongly suggests that sodium–ammonia solutions do *not* show superconductivity. Since none of the subsequent investigations succeeded in offering a simple explanation of the original observations, it is particularly unfortunate that no further report on this research has been made by Ogg himself.

A number of metals were tested for superconductivity in the temperature region below 1° K. by Alekseyevsky and Migunov (1947). Of these metals, silicon (0·073° K.), chromium (0·082° K.), antimony (0·152° K.), tungsten (0·070° K.), beryllium (0·064° K.) and rhenium (0·086° K.), none showed superconductivity at the lowest temperature (given in each case in brackets) reached. Uranium became superconductive at 1·3° K.; rhenium had previously been reported superconductive at 0·95° K. by Aschermann and Justi (1942), while a sample of uranium investigated by Shoenberg (1947) did not show superconductivity. Whereas superconductivity of rhenium appears unlikely owing to its position in the periodic table uranium may, according to modern conceptions concerning the metals above actinium, belong to the superconductive family (cf. I). More work is clearly indicated to elucidate its behaviour at low temperatures. Doubts have also arisen concerning the superconductive properties of titanium. It has been suggested (Webber and Reynolds 1948) that the observed drop in resistance between 3° K. and 1° K. may be due to the superconductivity of an alloy or a compound of titanium and that the pure metal will only become superconductive at a lower temperature. Titanium had previously been listed as a superconductor with a transition temperature of approximately 1·8° K.

As mentioned in I, H. London observed absorption of electromagnetic energy of high frequency in a superconductor, thereby demonstrating the penetration of a magnetic field into the skin of the superconductive metal. The depth D of this penetration and its dependence on temperature is of considerable importance for the theoretical interpretation of superconductivity. London (1947a) observed the Joule heating in an ellipsoid of superconductive tin exposed to electromagnetic radiation of $\sim$20 cm. wavelength. This work has been repeated and extended by Pippard (1947) using a new method. He makes the superconductor part of a resonating circuit and determines the H.F. resistivity from the selectivity of the resonator. Experiments carried out on tin with a wavelength of 25 cm. confirm the results of London, and the work has been extended to mercury using the same wavelength. In conjunction with observations on normal metals Pippard has worked out a theory of the skin resistance and applied this to the evaluation of D. This quantity may also be deduced from determinations of the skin reactance which was measured directly. The calculated curves for the variation of D with temperature show a fair but not wholly satisfactory agreement with the values obtained from the susceptibility of colloids (cf. I). The same is true for the comparison of the temperature dependence of D obtained from resistivity and from reactance. More recently Pippard (1948) has extended this research to a wavelength of 3·2 cm., employing the same technique. As was to be expected, the H.F. resistance increases with increasing frequency but the discrepancy between theory and experiment has become much worse.

Preliminary results with 3·2 cm. waves on lead have been reported by Bitter, Garrison, Halpern, Maxwell, Slater and Squire (1946) and it was announced by Slater (1948) that more recently the H.F. resistivity of lead and tin has been measured with 1·25 cm. waves down to 2° K. At this wavelength the H.F. resistivity does not tend to zero as the temperature is lowered but seems to approach a finite value.

Two methods used by Shoenberg and his co-workers have led to a more direct determination of D or at least of its dependence on temperature. Following up the work on the magnetic susceptibility of superconductive colloidal particles (cf. I) Désirant and Shoenberg (1948 a) have measured the magnetic induction in bundles of fine wires of tin and mercury. The magnetic induction in a superconductor is zero except for the thin surface layer into which a magnetic field can penetrate. A variation of the thickness D of this layer with temperature will therefore alter the effective cross-section of the specimen. Since this variation is only of the order of 10^{-5} cm. the total change in induction, even for fine wires, is rather small and requires a sensitive method of measurement. Here, too, the results are in fair agreement with the previous work, and the value of D at absolute zero is estimated to be $7{\cdot}6 \times 10^{-6}$ cm. The second method is that previously employed by Casimir (cf. I) in which the magnetic flux is measured inside a coil which is wound closely on a glass container filled with superconductive mercury. The result obtained by Casimir was most perplexing; no change of flux, i.e. no variation of D with temperature, was recorded. Later a possible source of error was discovered by Casimir himself, who pointed out that the sealed off glass capsule was immersed in liquid helium and that the change in vapour pressure with temperature might have caused a deformation roughly cancelling the change of magnetic flux. The experiment was therefore repeated (Laurmann and Shoenberg 1947) with the glass container not sealed off. Under these conditions results were obtained which are indeed in agreement with the temperature dependence of D yielded by other methods. Moreover, there are indications that in tin D depends slightly on the relative direction of the magnetic field and the crystallographic axes.

Summarizing all the work on the determination of D, which has taken up a large share of the whole research on superconductivity, the general variation of this quantity with temperature can now be regarded as well established. However, the D–T curve is not yet known with sufficient accuracy to permit unambiguous theoretical interpretation. The absolute value of D, too, is still a matter of some uncertainty.

In a series of systematic and detailed experiments Désirant and Shoenberg (1948 b) and Andrew (1948) have investigated the magnetization and resistance of superconductive cylinders in transverse magnetic fields. Applying Landau's model of the intermediate state (a system of branching laminae, cf. I) or a similar system of branching threads, the magnetization curves for wires of different diameter were worked out. For a cylinder of infinitely large diameter, penetration of a transverse field should occur at half the critical field ($\frac{1}{2}H_c$) and the magnetization should disappear at H_c. However, for any finite diameter the depth of penetration as well as the surface tension α between the normal and the superconductive phase has to be taken into account. It is convenient to express these corrections in terms of a quantity of the dimensions of a length (cf. I),

$$d' = d - D \quad [d = 8\pi\alpha/H_c^2].$$

The theoretical curves for radius $r = \infty$ and for a narrow wire are given in Figure 2(*a*). As can be seen, the increase in magnetization of the fine wire is maintained to fields higher than $\frac{1}{2}H_c$. Then a sudden drop takes place which leads to a gradual decline, the magnetization becoming zero at a field lower than H_c. Figure 2(*b*)

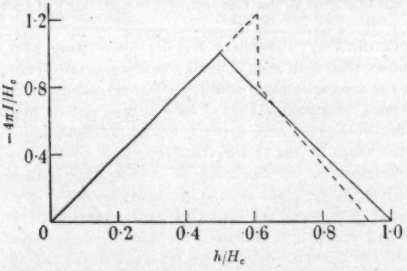


Figure 2*a*. Theoretical magnetization curves for a thick wire (——) and for a thin one (----).

gives the experimental values on a tin cylinder of $7{\cdot}5 \times 10^{-3}$ cm. diameter, showing that the features of the theoretical curve are well brought out by the observed data. In this case the electrical conductivity was measured simultaneously and it was noted that the rapid rise of the resistance coincided with the sudden

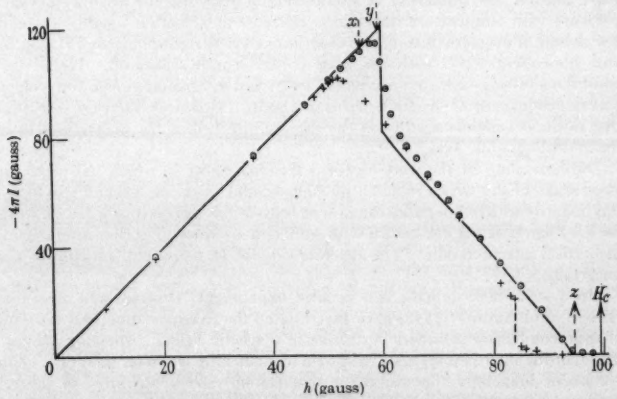


Figure 2*b*. Experimental magnetization curve of a thin wire of $7{\cdot}5 \times 10^{-3}$ cm. diameter. (*Désirant and Shoenberg.*)

drop in magnetization. This work thus explains an obscure phenomenon on which a number of papers have been published: the fact that the return of the resistance does not occur at $\frac{1}{2}H_c$. Apart from this most important result this careful work has provided general insight into the case of transverse magnetization.

Meshkovsky and Shalnikov (1947) have continued the work on the structure of the intermediate state in a superconductive sphere. Again a bismuth probe was used in a slit between two lead hemispheres. The results give evidence of an inhomogeneous and probably laminar intermediate state, but the theory is not obeyed in detail. It was shown some time ago (Mendelssohn and Pontius 1936) that the same sample of superconductive metal will show quite different magnetic properties whether it is in the form of a sphere or a short cylinder. In the first case the sample passes through the intermediate state and no residual flux is left once the external field is reduced to zero. The cylinder, on the other hand, will retain some of the flux; there exist then macroscopic normal and superconductive regions and the latter must carry persistent currents retaining the residual flux. Alekseyevsky (1946) has investigated in detail the distribution of these regions in cylindrical discs of superconductive lead. In preliminary experiments iron filings were used to render the lines of force visible. The main work, however, was carried out with a small search coil. The distribution varies conspicuously with the strength of the external field applied. In the case of a small field the flux only penetrates near the rim of the disc, the centre remaining superconductive, but for fields approaching H_c the residual flux is strongest in the centre.

Following up earlier observations on lead, MacDonald and Mendelssohn (1948b) have investigated the transition curve of the resistance in lead, mercury and tin in longitudinal magnetic fields. They found that in all cases the normal value of resistivity was only reached at a field somewhat higher than H_c. The effect was most pronounced in lead, which is probably due to the fact that this metal has the highest H_c values. It was noted that although the specimens used were very pure and were probably single crystals, no hysteresis effects were observed. Hysteresis did, however, occur (chiefly to mercury) when the ends of the wires, carrying the leads, were made of spherical shape. It appears evident that hysteresis effects are being produced by geometrical factors which, curiously enough, also have the effect of narrowing down the transition region. Thus there seem to be, very pronounced in lead, indications in the normal resistance when the superconductive state is approached in an external magnetic field. Another sign of a deviation from the normal behaviour of the metal is reported by Casimir and Rademakers (1947) who confirmed earlier indications that there occurred a sharp change in the thermo-electric force of tin when the transition point (in zero field) was approached. This deviation began to be noticeable about 0·2° K. above the temperature at which the resistance drops to zero.

The suitability of superconductors for use as sensitive bolometers has been further investigated by Andrews and his group (Andrews, Milton and DeSorbo 1946, Andrews 1947). In the course of this work the transition points for a number of hydrides and nitrides have been determined (Horn and Ziegler 1947) and the transmission of light through lead films was studied (Wexler 1946). In the latter experiments visible light from an incandescent filament lamp and from a high pressure mercury arc was observed through a thin lead film which was then cooled to superconductivity. In neither case did a change in the absorption occur.

In view of the lack of a generally accepted theory, superconductivity has always invited semi-empirical speculations. Sienko and Ogg (1947) have tried to connect the critical fields H_c and the temperature T by the relation

$H_c = A(T_s^{\frac{1}{2}} - T^{\frac{1}{2}})$, in which A is a constant and T_s the transition temperature in zero field. However, Stout (1947), as well as Daunt (1947), have pointed out that the experimental values cannot be represented in this way whereas they fit fairly well the old formula $H_c = B(T_s^2 - T^2)$, in which B varies slightly from metal to metal. De Launay and Dolecek (1947) have tried to find a relation between the Debye characteristic temperature and the transition point of metals. While it seems to be possible to group together a number of metals in a curve in this diagram, others do not conform to this rule. This is not really surprising since earlier work has shown that for the establishment of such regularities at least two parameters are required. The present authors predict superconductivity in cerium, praseodymium, neodymium and actinium, but believe that scandium and yttrium are unlikely to become superconductive. It has been pointed out (Dixit 1948) that according to this relation superconductivity should not occur in any metal at temperatures appreciably below 0·7° K.

The theoretical interpretations of superconductivity have followed as in previous years two rather different lines of approach. Firstly the phenomenological treatment developed by F. and H. London has been extended, and secondly a number of attempts have been made at elucidating the operative mechanism of superconductivity. von Laue has dealt in great detail with the theory of the electrodynamical phenomena (v. Laue 1947a, b, c). In the last of these papers the theory had been extended to non-cubic crystals. The work constitutes on the whole an amplification of the papers by the same author already discussed in I. In this context a criticism by Broer (1947) of the interpretation of the gyromagnetic effect in superconductors should be mentioned. He shows that the observation of the gyromagnetic constant as equal to unity does not permit any conclusion as to the mechanism of superconductivity but that it is a necessary consequence of the conservation of angular momentum. Similarly the existence of the penetration depth in a superconductor could be predicted without additional assumptions. Philbert (1946) has applied the London equation to the case of thin wires and concludes that the latter should have a higher magnetic and lower current threshold than the bulk material; this had already been postulated by various authors and verified by direct experiments. An interesting suggestion has been made by Ginsburg (1947) who, on the basis of the London theory, discusses the dependence of the penetration depth on the external magnetic field. When the latter approaches the critical value the number of "superconductive electrons" is decreased and the equations for the current and the field become non-linear. Such an increase in penetration depth near the threshold curve would also appear to be a necessary consequence of Heisenberg's theory, and a decision by direct experiment appears to be feasible. The only attempt so far (Désirant and Shoenberg 1948a) has, however, yielded an inconclusive result.

A good deal of attention has been attracted by a theory on the mechanism of superconductivity proposed by Heisenberg (1947). This model assumes the formation of something like an electron crystal which is produced by the long range Coulomb interaction. The electrons affected are those of maximum energy and there takes place a kind of condensation phenomenon on the surface of the Fermi sphere. It would appear that this condensation in momentum space, which is particularly emphasized in a second paper (Heisenberg 1948), is of greater significance than the postulated ordering in space. At absolute zero the surface of the Fermi sphere is completely covered by the "electron condensate"

a condition which seems to require a falling off at low temperatures of the critical field to zero. At low but finite temperatures part of the condensed layer on the momentum sphere will have evaporated, leaving a hole. Owing to the Coulomb exchange force such a single hole is energetically preferable to a number of smaller holes of equal size, but it leads to a finite current density. Thus, contrary to the assumptions of the London electrodynamics, the lowest energy state in this model carries a current. Koppe (1947, 1948) has shown that Heisenberg's theory permits the evaluation of the temperature function of the specific heat and the order of magnitude of transition temperatures which are in agreement with the experimental facts. A particularly interesting prediction of this theory is that, apart from the ferromagnetic ones, all metals must become superconductors. Born and Cheng (1948) on the other hand have found an empirical rule between lattice structure and occurrence of superconductivity. They have drawn attention to the fact that superconductors are metals in which the Fermi sphere almost coincides with the corners of a Brillouin zone, whereas in non-superconductive metals, like the alkalis, the radius of the sphere and the distance of the corners from the centre are rather different. The authors interpret this relation as indicating that superconductivity is due to a redistribution of electrons from the corners of the zone on to the top of the Fermi surface. The theory thus combines ionic forces, responsible for a favourable relation between Fermi surface and Brillouin zone, with electron interaction, to which the redistribution is ascribed. Calculation on the basis of such a model shows that the state of lowest energy may be one in which only some of the corners give off electrons to the Fermi surface. Such an unsymmetric occupation represents, however, a state with a spontaneous current. The conception of a lowest state which is not current-free and which is a characteristic feature of both theories has been attacked by F. London (1948), who points out that there exists no experimental evidence for such a state. London, while not presenting a rigorous theory, suggests, like Heisenberg, that superconductivity is due to exchange interaction associated with the Coulomb field, but he arrives at somewhat different conclusions. According to his view ferromagnetism and superconductivity are the opposite limiting cases of this interaction competing with the zero point energy. Whereas in the former case the result is a parallel orientation of the electron spins, in the latter it is an ordering of the momenta in a state of vanishing total spin.

The general impression one derives is that the three theories may not differ quite so much as appears at first sight from the critical attitude taken up by the respective authors. Common to all is the important rôle assigned to the Coulomb exchange interaction as the operative agent for the formation of the superconductive state. The space order postulated by the Heisenberg theory, which is particularly criticized by F. London, does not seem to be more than an incidental, and possibly unnecessary, feature of the former. The empirical rule discovered by Born and Cheng gives a welcome support to the experimental fact that only certain related groups of metals have been found to be superconductive. Their interpretation of its physical significance, however, may prove more a correction of the Heisenberg theory than a contradiction. Thus, while these theories are still somewhat vague as far as detailed treatment is concerned, the field of speculation appears to have been narrowed down remarkably when comparing it with the attempts of previous years. Although there are now positive indications as to the kind of process responsible for the phenomenon of superconductivity,

and in spite of the great amount of available data, one may suspect that certain important experimental facts are still missing. This is emphasized by the controversy as to whether the ground state carries a current or not.

§4. LIQUID HELIUM

The peculiar properties of liquid helium, which have already been the subject of a previous Report (Jones 1939) have led to a great deal of experimental as well as theoretical work. The discovery of the thermo-mechanical ("fountain") effect (Allen and Jones 1938) had given rise to the idea (Tisza 1938) that liquid helium II (below 2·19° K.) might be considered as an interpenetrating mixture of a normal and a superfluid liquid. This conception gained strong experimental support when it was found (Daunt and Mendelssohn 1939b) that helium which was "filtered" through a fine powder cooled. This mechano-caloric effect, which is the reverse of the above-mentioned phenomenon, thus shows that it is possible to separate by capillary flow the two constituents of the mixture and that the superfluid part has a lower entropy than the liquid as a whole. Accurate determinations of the entropy loss were interrupted at the beginning of the war but were two years later carried out by Kapitza (1941b) who showed that the liquid passing through a narrow slit lost *all* its entropy. This conclusion rests of course on a knowledge of the specific heats of liquid helium extrapolated to absolute zero. The experiments of Keesom and Westermijze (1941), which have yielded between 0·6° K. and 1·5° K. a T^6 function of the specific heat, indicate indeed that the entropy of the superfluid liquid must be either zero or very small. The model conception of two interpenetrating liquids, one of which exhibits the phenomenon of potential flow, is further supported by a number of experiments carried out by Kapitza (1941a) which were designed to elucidate the mechanism of the very high heat conduction of helium II. In one of them a glass tube filled with the liquid was used which contained a concentric rod. This rod could be rotated up to 1,900 rev/min. Comparison of the heat conduction without and with rotation showed that in the latter case the heat flow was much reduced (in one case to 1/100). The same happened when the liquid was made to flow along the tube either with or against the heat current or if it was disturbed in any other way. All this pointed to some sort of a convection process underlying the high heat flow. This suggestion had already been made by various authors. Kapitza also made experiments on the heat flow from a thermally insulated bulb which was fitted with a single capillary as outlet. The bulb could rotate freely in the main helium bath (Figure 3), and it did so when the heating was on. The conclusion to be drawn from this observation is that the capillary acts as a jet emitting a stream of helium, the reaction of which drives the bulb forward. This was proved by fitting a disc to the bulb; when this covered the outlet of the capillary the motion was stopped. Separate experiments were made on the "topography" of the jet. In the stationary state of constant heat flow along the capillary there must therefore exist a counterflow of helium into the bulb which carries no momentum. It appears that these two opposing flows are interpenetrating although it would seem that the possibility of the counterflow being limited to a layer along the inner wall of the capillary cannot be rejected. Quantitative measurements of the heat transported in this convection show that the entropy of the liquid in the counter-current must be zero or at least very low. The earlier experiments by Keesom, Saris and Meyer (1940) showing a complex

relation between heat flow, temperature difference and the dimensions of the experimental arrangement must clearly also be interpreted in this way.

Experiments similar to those of Allen and Reekie and Kapitza correlating heat conductivity and thermo-mechanical effect with the applied heat current have been carried out by Keesom and Duykaerts (1947). These measurements, in which a narrow slit formed the connecting link between the two volumes of helium, have been extended by Mellink (1947) and Meyer and Mellink (1947). In these experiments the change of the phenomena associated with the heat flow was systematically investigated as the width of the slit was decreased to $0{\cdot}1\,\mu$. For wider slits ($\sim 10\,\mu$) the temperature dependence of both the heat conductivity

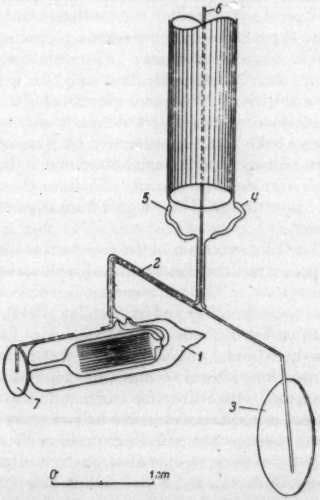


Figure 3. Kapitza's experiment on the mechanism of heat conduction in liquid helium II.
1, Bulb with heater ; 2, crossbar ; 3, counterweight ; 4, 5, current leads ;
6, quartz fibre ; 7, disc.

and the thermo-mechanical effect shows the previously established shape of a smooth curve with a maximum in the neighbourhood of $1{\cdot}8°$ K. There is a gradual drop to zero at the λ-point. However, the character changes completely as the slit is made narrower, and in the case of the smallest slits heat conduction and thermo-mechanical effect rise with increasing temperature in monotonic functions with a sudden fall at the λ-point. The mechanism of the heat flow also seems to have changed. In capillaries and wider slits the heat flow changes with the cube root of the temperature difference, but in the narrow slits the heat current is directly proportional to ΔT. In agreement with Kapitza's experiment, H. London's equation $dp/dT = \rho \Delta S$ (cf. I) is obeyed, p being the pressure, ρ the density and S the entropy. However, this relation breaks down above a certain

flow velocity through the slit. Once this velocity is exceeded, the process becomes irreversible and overheating occurs, which results in thermal relaxations. The existence of such a critical velocity was also observed by Daunt and Mendelssohn (1939 a) in observations on helium films. It is therefore of interest to note that the critical velocity observed in these experiments and its dependence on temperature was, within the limits of accuracy, the same as that in the films. It should be mentioned, however, that some observations of Kapitza (1941 b) indicate a critical velocity about five times larger. Another experiment supporting the two-fluid conception has been carried out by Andronikashvilli (1946) who has measured the inertia of a vessel containing a stack of circular metal plates oscillating in liquid helium. As the temperature is decreased below the λ-point an increasing part of the liquid will cease to take part in the oscillation and the moment of inertia grows smaller. The experiment thus provides a measure of the change of the superfluid constituent with temperature. The result obtained confirms the two-fluid conceptions derived from the flow and film experiments mentioned above and also from the viscosity measurements using an oscillating disc; all these point to an increase of the superfluid part with lowering temperature. The latter experiments have been supplemented by Keesom and Keesom (1941) by more accurate determinations in the neighbourhood of the λ-point. It appears now that at no temperature does the viscosity of helium II exceed that of helium I, and that there is no discontinuity in the temperature dependence of the viscosity at the λ-point.

One of the most striking successes of the two-fluid model is the experimental verification of the prediction that heat impulses will travel through the liquid without suffering dissipation. This phenomenon, which is now generally referred to as "second sound", was first observed by Peshkov (1944), whose work has since been repeated and extended by Lane and others (Lane, Fairbank and Fairbank 1947, Fairbank, Fairbank and Lane 1947). In these experiments an electric heater of small heat capacity was used as source and a (phosphor-bronze) resistance thermometer as detector. An alternating current fed into the heater set up temperature oscillations in a column of liquid helium. By changing the relative position of source and thermometer standing waves of the second sound can be measured. The velocity of these temperature waves was found to rise from zero at the λ-point to about 20 m/sec. at 1·8° K. Below this temperature the second sound velocity is almost independent of temperature, showing, however, a slight tendency to decrease below 1·7° K. In the work of Lane and his collaborators a microphone suspended above the liquid level was used as detector. The thermal vibrations produce a periodic evaporation on the liquid surface which is recorded as ordinary sound. Pellam (1948) and Osborne (1948) have both measured second sound by directly timing the velocity of a single pulse. The former has for the first time observed attenuation in second sound, which was found to be small, showing an increase with temperature. Pellam and Squire (1947) have repeated the earlier determinations of ordinary sound in liquid helium (Findlay, Pitt, Grayson-Smith and Wilhelm 1939), which had been carried out at 1·3 Mc/s. with a frequency of 15 Mc/s. The new results fully confirm the earlier work and also add determinations of the attenuation. The attenuation of helium I shows a minimum at about 3° K.; there follows a sharp drop at the λ-point and in helium II a further rise is observed as the temperature is lowered.

A number of experiments have been made on the helium II film. Daunt and Mendelssohn (1939b) have used an arrangement in which heat could be supplied to the inside of a small Dewar vessel suspended in a bath of liquid helium. Without heating, the liquid levels inside and outside adjust themselves to the same height. However, when heat is supplied the inside level rises in spite of the increased evaporation. This phenomenon is analogous to the thermo-mechanical effect, except that in this case the connecting link is not a capillary but the film. This result emphasizes the fact that the film offers a convenient method of making observations on the superflow. Further experiments with similar arrangements showed that the rate at which the level difference was established was the same (Mendelssohn 1947, details unpublished) as the critical rate of isothermal film transfer from a higher to a lower level. In the latter type of experiment the film flow is always limited by this critical rate which, on the other hand, is independent of the level difference. In order to determine the difference in gravitational potential at the ends of the film, Daunt and Mendelssohn (1946b) employed a device in which the flow occurred at less than the critical rate. Under these conditions it was found that within the limit of accuracy the potential difference at the ends was zero; it was less than 15 times smaller than the level difference required to accelerate the helium to the observed velocity by free fall. Atkins (1948) has found a film flow of a higher rate than that previously observed; this also depended on the liquid level when, instead of a beaker, a bulb with a narrow neck was used. He also has observed a higher rate of transfer if the film flows under the influence of a temperature difference. The last result appears to be in disagreement with the measurements of Daunt and Mendelssohn mentioned above as well as with an observation by Meyer and Mellink (1947, p. 204). These facts clearly indicate that our knowledge of the phenomenology of film flow is still rather incomplete.* This holds in particular for the formation of the film, about which little is known. Daunt and Mendelssohn used the high heat transport initiated by the film as an indicator for its formation. They found (Mendelssohn 1947b) that the rate at which the film is deposited from the gas phase is less than 10% of the rate at which it rises out of the liquid. Experiments by Kistemaker (1947) on the adsorption of helium on glass show a large number of layers ($\sim$30) not only in the helium II region but also ($\sim$15) above the λ-point. The 30 layers observed in this case are rather less than the thickness of the transfer film ($\sim$100 layers), and the question arises whether the presence of the bulk liquid is necessary for its formation. Brown and Mendelssohn (1947) have separated the helium film from the bulk liquid out of which it is formed. This was done by using between two volumes a connecting tube, which was tightly packed with a great number of wires. Such a tube acts as a "filter" which will allow the passage of the frictionless film but stops the viscous flow of the gas phase. Using such an arrangement these authors found that the isolated film is in equilibrium with a gas phase having the same vapour pressure as the liquid.

It is now clear that the most important step in the interpretation of the λ-phenomenon was the proposal by F. London (already discussed, Jones 1939) to consider it as the Bose–Einstein condensation phenomenon. It also appears that the theory does not owe its significance to the success of the particular model but because it has opened up an entirely new line of approach. The idea of a

* Recent experiments (Bowers and Mendelssohn 1949) suggest that the high transfer rates are in fact due to contamination of the transfer surface with condensed air.

"phase" which is separated from the normal liquid not in coordinate but in momentum space at once introduced the concept of two interpenetrating fluids which has provided a most useful working hypothesis for the experimentalists. As such it has been eminently successful, and this success has in turn influenced further theoretical attempts. The first attempt at making predictions on the bases of a phenomenological model of this kind (Tisza 1947a, cf. Jones 1939) has met with signal success. It led among other things to the prediction of the "second sound". However, this short-cut, while extremely useful to a degree, is severely limited in its application, and in order to fit the observed facts such essential features as the zero momentum of the Bose–Einstein condensate have to be abandoned (Mendelssohn 1947a). The position has been summarized by F. London (1947), who regards the rigorous treatment of the Bose–Einstein model as presenting "apparently hopeless details" and can therefore see some justification for Tisza's method. Attempts have indeed been made to work out the properties of a Bose–Einstein model for a real gas or liquid. Schiff (1941) has calculated the partition function for a non-ideal quantum gas, neglecting inter-particle forces except for binary collisions. Application of this treatment to the Bose–Einstein model of helium fails to account for the discontinuity in the specific heat, and the author concludes that the gas approximation is unsatisfactory. Goldstein (1946) on the other hand, assessing the properties of Bose–Einstein fluids in which the attractive forces are averaged over the whole volume, expects a first order transition. Introduction of an equally averaged repulsive force then results in a λ-point. An attempt to account for the phenomenon of superfluidity by considering a non-ideal Bose–Einstein gas has been made by Bogolubov (1947), who also considers the use of a semi-phenomenological treatment as a necessary simplification. Bijl, de Boer and Michels (1941) have suggested an explanation for the helium film due to the existence of "condensed" atoms. The kinetic zero point energy of such a system (due to indeterminacy) depends on the inverse square of the linear dimensions, and this leads to an energy difference between the bulk liquid and the thin layer rising up on a solid surface which dips into the liquid. Equating this extra energy with the work done against gravity, the authors obtain a film thickness of the same order ($\sim 3 \times 10^{-6}$ cm.) as that found experimentally (Daunt and Mendelssohn 1939a).

An interesting suggestion with reference to the helium problem has been made by Gentile (1941) who has proposed a new type of quantum statistics, intermediate between those of Bose–Einstein and Fermi–Dirac. Whereas in the former the maximum number of particles per phase cell is either one or ∞, the new statistics proposes to change the number ∞ to n, this being the number of particles in the system. Gentile himself (1942), and later other Italian workers (Salvetti 1941, Caldirola 1943), have attempted to apply the intermediate statistics to liquid helium, trying to account in this way for the anomaly in the specific heat as well as for the transport phenomena. A number of papers (Sommerfeld 1942, Caldirola 1946, Schubert 1946, Borsellino 1947) have since been published discussing the applicability of the Gentile method.

Landau (1941, 1944, 1947) has developed a theory of superfluidity, the salient feature of which is the introduction of the "roton" conception. The excitations in liquid helium are supposed to consist of phonons, i.e. Debye heat capacity of longitudinal waves and of rotons, "elementary excitations of the vortex spectrum". This quantization of liquid motion has been criticized by F. London (1947) as having no well-defined foundation and also as not leading

to the macroscopic model of two interpenetrating fluids which Landau postulates. This, the phenomenological part of Landau's theory, deals with the hydrodynamics of such a mixture and is in many respects similar to Tisza's model. Landau's model, too, leads to the phenomenon of second sound, but whereas Tisza's theory expects the second sound velocity to drop off to zero as absolute zero is approached, Landau's theory (Lifshitz 1944) predicts a rise to $1/\sqrt{3}$ of the velocity of ordinary sound. As mentioned above, the earlier experiments showed indeed a maximum in the velocity at $\sim 1{\cdot}7°$ K. However, a recent extension to $1{\cdot}03°$ K. (Peshkov 1948) indicates a further rise below $1{\cdot}1°$ K. While the final decision has to be left to experiments at still lower temperatures, the existing evidence seems at present to favour Landau's theory.

A discussion of the theoretical interpretation of second sound can be derived simply from the model of interpenetrating fluids without specialized assumptions. Properties of second sound and its relation to heat transport have been discussed

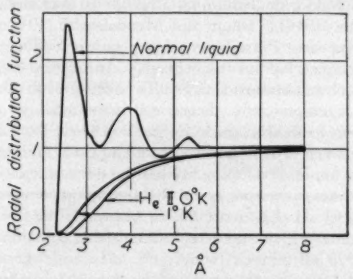


Figure 4. Radial distribution function of a normal liquid (argon) and helium II at 0° K. and 1° K. (*Born and Green.*)

in a number of papers by Band and Meyer (Band and Meyer 1948 a, b, Meyer and Band 1948) which are based on a two-fluid conception. These authors introduce the conception of a relaxation time in a wave equation for second sound. They derive a relation between the velocities of ordinary and second sound. In this treatment the heat conduction of liquid helium is connected with the second sound in the same manner as normal heat conductivity is related by the Debye theory to ordinary sound.

A theory of liquid helium not based on analogies with gas models but considered as a limiting case of the theory of liquids has been given by Born and Green (1947 a, b). In such a kinetic theory of a quantum liquid the states of angular momentum $l=2$, 4, etc. give no contribution at very low temperatures, and this results in a radial distribution function very different from that of a normal liquid (Figure 4). The temperature at which the last non-vanishing state of angular momentum begins to disappear is very close to the experimentally determined temperature of the λ-point. It is a feature of the theory that the classical conceptions of temperature and pressure are no longer applicable. Distinction is made for instance between the hydrodynamic and the thermodynamic pressure, which in helium II have different values. (The fact has, however, been disputed by Yvon 1948.) The theory explains the transport phenomena as mass flow which is produced by a heat flow normal to the direction of the former. The film

transport, for instance, is thought to be due to evaporation from the surface of the film.

As already mentioned in I, Daunt and Mendelssohn (1942) have drawn attention to the analogous behaviour of liquid helium II and superconductors. On a purely phenomenological basis a model has been proposed (Mendelssohn 1945) which extends the two-fluid conception to both phenomena. A condensation in momentum space is assumed which does not, however, ascribe zero momentum to the condensed phase. The two phenomena are thus considered not as special subjects but as manifestations of a fourth state of aggregation in which frictionless transport is closely associated with zero entropy without order in coordinate space. A number of attempts have since been made to account for this analogy. F. London (1945) has taken up the problem of frictionless transport and proposed the general connection between the rate R of transport and the number n of superfluid particles $R<n\hbar/2\pi$. A similar relation had already been obtained for the helium film by Bijl, de Boer and Michels (1941) and by Gogate and Rai (1944). Daunt and Mendelssohn (1946 a) have pointed out that this relation appears to be a direct consequence of a diffusion of the superfluid constituent under its zero point momentum. An experiment designed to prove the existence of such a mechanism has given a positive result (Daunt and Mendelssohn 1946 b). Attempts at a theoretical interpretation of this frictionless state of aggregation have been made by Band (1946 a, b). Theoretical reasons for the analogy between superconductivity and liquid helium have also been given by Green (1948) who, in developing further the general kinetic theory of liquids of Born–Green, comes to the conclusion that the two phenomena can be described by the same theory. H. London (1947 b) has made the interesting suggestion that possibly in analogy to superconductivity the equation of potential flow in helium d curl $v/dt=0$ may be reduced to curl $v_s=0$ where v_s is the velocity of superfluid flow. In a recent paper on the thermodynamics of superconductivity Heisenberg (1948) too expresses the opinion that superconductivity and liquid helium form a superfluid state of aggregation which takes an equal place with the solid state.

§5. HELIUM 3

In view of the peculiar properties of liquid helium, special interest has been attached to the possibility of obtaining the isotope with the mass 3 in sufficient quantities to make experiments at low temperatures. ^{3}He has an abundance ratio of $1{\cdot}3\times10^{-6}$ in atmospheric helium and only $1{\cdot}6\times10^{-7}$ in helium derived from natural gas. When working with dilute solutions of ^{3}He in ^{4}He it is therefore necessary to use a sensitive mass spectrograph for the analysis of samples. In this the American workers had the advantage of the excellent mass spectrographic technique of Aldrich and Nier. The first experiment on the separation of ^{3}He by a low temperature method was carried out in Yale (Fairbank, Lane, Aldrich and Nier 1947). Atmospheric helium was condensed and the vapour was sampled at different temperatures between 4·2° and 2·2° K. and subsequently analysed. From the fact that the relative concentration of ^{3}He increased with decreasing temperature the authors could conclude that if pure ^{3}He exists as a liquid its boiling point must be lower than that of ^{4}He. Thus fractionated distillation appears a possible method for low temperature separation of the isotopes. A different way of separation was opened when it was discovered (Daunt, Probst, Johnston, Aldrich and Nier 1947) that ^{3}He did not take part in

film flow or in flow through narrow slits (Daunt, Probst and Johnston 1947). The same approach was further pursued by Lane, Fairbank, Aldrich and Nier (1948), who separated ^{3}He by means of the thermo-mechanical effect. In this work a tube containing the unseparated liquid was heated at one end and this resulted in an increase in concentration of ^{3}He at the cold end.

It has sometimes been assumed (cf. Franck 1946) that the failure of ^{3}He to take part in superflow is a proof of London's Bose–Einstein model of liquid helium II because the isotope with mass 3 must obey Fermi–Dirac statistics. However, it has been pointed out by Landau and Pomeranchuk (1948) that in small concentration foreign particles like electrons or ^{3}He atoms cannot take part in superflow, and these authors therefore think that it cannot be deduced from the experimental evidence whether or not ^{3}He in liquid form will have a λ-point.

A rather disturbing fact was noted by Fairbank, Lane, Aldrich and Nier (1948) continuing their experiments in which the gas phase above condensed atmospheric helium was analysed: while above the λ-point the ^{3}He concentration increased with decreasing temperature, this relation was found to be reversed below the λ-point. Theories of solutions of ^{3}He in ^{4}He have been made by Band (1948) and by London and Rice (1948). The latter authors try to account for the peculiar behaviour of ^{3}He below the λ-point by assuming the formation of "globules" of superfluid ^{4}He round the ^{3}He atoms. It should be noted, however, that Rollin and Hatton (1948) in Oxford, carrying out similar experiments with a liquid which had been enriched with ^{3}He to 0·15%, found a rather different result. They had become suspicious that below the λ-point the ordinary sampling methods might be largely falsified by evaporation from the helium film and had accordingly designed their experiments so as to avoid this effect as much as possible. Evaporating helium at 1·3° K., they found indeed a ^{3}He concentration of 1·5% in the vapour. It is believed that but for unavoidable film flow the enrichment factor would have been even greater than 10-fold. It is thus possible that the earlier results are obscured by evaporation of the film although it has of course to be remembered that the different concentration used in the two sets of experiments may influence results.

London and Rice also state that owing to its high zero point energy "it is almost certain that pure ^{3}He cannot existi n a liquid phase at any temperature". This view is shared by Tisza (1947 b). De Boer (1948 a) on the other hand has predicted that the critical temperature of pure liquid ^{3}He should be between 3·1° K. and 3·5° K. and the critical pressure between 0·95 and 1·3 atmospheres. At 2° K. the vapour pressure of ^{3}He should be six times as high as that of ^{4}He. These data were calculated by means of a reduced equation of state (de Boer 1948 b) in which quantum effects are accounted for by the introduction of a quantum parameter containing $\boldsymbol{h}$ and the mass of the particle. These predictions of de Boer have indeed been verified to an astonishing degree by the recent experiments of Sydoriak, Grilly and Hammel (1949). These authors, working at the Los Alamos Laboratory, have condensed pure ^{3}He, obtained by β-decay from pure tritium, in a tube dipping into a bath of ordinary liquid helium. The apparatus did not permit visual observation, but the indications are that the condensed phase was liquid rather than solid. The critical point was found at 3·34° K. at a critical pressure of 1·2 atmospheres (87·5 cm. Hg). The vapour pressure at 2° K. was measured as ~14·5 cm. Hg, which is ~6 times greater than that of ^{4}He at this temperature (~2·35 cm. Hg).

REFERENCES

ALEKSEYEVSKY, N., 1946, *J. Phys. U.S.S.R.*, **10**, 360.
ALEKSEYEVSKY, N., and MIGUNOV, L., 1947, *J. Phys. U.S.S.R.*, **11**, 95.
ALLEN, J. F., and JONES, H., 1938, *Nature, Lond.*, **141**, 243.
ANDREW, E. R., 1948, *Proc. Roy. Soc.* A, **194**, 80.
ANDREWS, D. H., 1947, *Report of Cambridge Conference on Low Temperatures* (London : Physical Society), p. 56.
ANDREWS, D. H., MILTON, R. M., and DESORBO, W., 1946, *J. Opt. Soc. Amer.*, **36**, 518.
ANDRONIKASHVILLI, E., 1946, *J. Phys. U.S.S.R.*, **10**, 201.
ASCHERMANN, G., and JUSTI, E., 1942, *Phys. Z.*, **43**, 207.
ASTON, J. G., and SZASZ, G. J., 1947, *J. Chem. Phys.*, **15**, 560.
ATKINS, K. R., 1948, *Nature, Lond.*, **161**, 925.
BAND, W., 1946 a, *Proc. Phys. Soc.*, **58**, 302; 1946 b, *Proc. Camb. Phil. Soc.*, **42**, 311; 1948, *J. Chem. Phys.*, **16**, 343.
BAND, W., and MEYER, L., 1948 a, *Phys. Rev.*, **73**, 226; 1948 b, *Ibid.*, **74**, 386.
BIJL, A., DE BOER, J., and MICHELS, A., 1941, *Physica*, **8**, 655.
BIRCH, A. J., and MACDONALD, D. K. C., 1947, *Trans. Faraday Soc.*, **43**, 792.
BITTER, F., GARRISON, J. B., HALPERN, J., MAXWELL, E., SLATER, J. C., and SQUIRE, C. F., 1946, *Phys. Rev.*, **70**, 97.
BLACKMAN, M., EGERTON, A., and TRUTER, E. V., 1948, *Proc. Roy. Soc.* A, **194**, 147.
BLANCHARD, F. R., and BITTNER, H. W., 1942, *Rev. Sci. Instrum.*, **13**, 394.
DE BOER, J., 1948 a, *Proc. Phys. Soc.*, **61**, 578; 1948 b, *Physica*, **14**, 139.
BOGOLUBOV, N., 1947, *J. Phys. U.S.S.R.*, **11**, 23.
BOORSE, J. A., COOK, D. B., PONTIUS, R. B., and ZEMANSKY, M. W., 1946, *Phys. Rev.*, **70**, 92.
BORN, M., and CHENG, K. C., 1948, *Nature, Lond.*, **161**, 968, 1017.
BORN, M., and GREEN, H. S., 1947 a, *Nature, Lond.*, **159**, 738; 1947 b, *Proc. Roy. Soc.* A, **191**, 168.
BORSELLINO, A., 1947, *Nuovo Cim.*, **4**, 1.
BOWERS, R., and MENDELSSOHN, K., 1949, *Nature, Lond.*, **163**, 870.
BROER, L. J. F., 1947, *Physica*, **13**, 473.
BROWN, J. B., and MENDELSSOHN, K., 1947, *Nature, Lond.*, **160**, 670.
BURKHARDT, G., 1948, *Z. Naturforsch.* **3a**, 46.
CALDIROLA, P., 1943, *Nuovo Cim.*, **1**, 205; 1946, *Riv. Nuovo Cim.*, **2**, 25.
CASIMIR, H. B. G., and RADEMAKERS, A., 1947, *Physica*, **13**, 33.
COLLINS, S. C., 1947, *Rev. Sci. Instrum.*, **18**, 157.
DAUNT, J. G., 1947, *Phys. Rev.*, **72**, 89.
DAUNT, J. G., DÉSIRANT, M., MENDELSSOHN, K., and BIRCH, A. J., 1946, *Phys. Rev.*, **70**, 219.
DAUNT, J. G., and MENDELSSOHN, K., 1939 a, *Proc. Roy. Soc.* A, **170**, 423; 1939 b, *Nature, Lond.*, **143**, 719; 1942, *Ibid.*, **150**, 604; 1946 a, *Phys. Rev.*, **69**, 126; 1946 b, *Nature, Lond.*, **157**, 839.
DAUNT, J. G., PROBST, R. E., and JOHNSTON, H. L., 1947, *J. Chem. Phys.*, **15**, 759.
DAUNT, J. G., PROBST, R. E., JOHNSTON, H. L., ALDRICH, L. T., and NIER, A. O., 1947, *Phys. Rev.*, **72**, 502.
DÉSIRANT, M., and SHOENBERG, D., 1948 a, *Proc. Phys. Soc.*, **60**, 413; 1948 b, *Proc. Roy. Soc.* A, **194**, 63.
DESORBO, W., MILTON, R. M., and ANDREWS, D. H., 1946, *Chem. Rev.*, **39**, 403.
DINGLE, R. B., 1948, *Proc. Phys. Soc.*, **61**, 9.
DIXIT, K. R., 1948, *Phys. Rev.*, **73**, 182.
FAIRBANK, H. A., 1946, *Rev. Sci. Instrum.*, **17**, 473.
FAIRBANK, H. A., LANE, C. T., ALDRICH, L. T., and NIER, A. O., 1947, *Phys. Rev.*, **71**, 911; 1948, *Ibid.*, **73**, 729.
FAIRBANK, W. M., FAIRBANK, H. A., and LANE, C. T., 1947, *Phys. Rev.*, **72**, 645.
FINDLAY, J. C., PITT, A., GRAYSON SMITH, H., and WILHELM, J. O., 1939, *Phys. Rev.*, **56**, 122.
FRANCK, J., 1946, *Phys. Rev.*, **70**, 561.
GENTILE, G., 1941, *Ric. Sci.*, **12**, 341; 1942, *Nuovo Cim.*, **19**, 109.
GIBNEY, R. B., and PEARSON, G. L., 1947, *Phys. Rev.*, **72**, 76.
GINLOTTO, L., and GIGLI, A., 1947, *Phys. Rev.*, **71**, 211.
GINSBURG, V., 1947, *J. Phys. U.S.S.R.*, **11**, 93.

GOGATE, D. V., and PATHAK, P. D., 1947, *Proc. Phys. Soc.*, **59**, 457.
GOGATE, D. V., and RAI, R. N., 1944, *Nature, Lond.*, **153**, 342.
GOLDSTEIN, L., 1946, *J. Chem. Phys.*, **14**, 276.
GREEN, H. S., 1948, *Proc. Roy. Soc.* A, **194**, 244.
DE HAAS, W. J., and DE BOER, J., 1934, *Physica*, **1**, 609.
HEISENBERG, W., 1947, *Z. Naturforsch.*, **2a**, 185; 1948, *Ann. Phys., Lpz.*, **3**, 289.
HILSCH, R., 1947, *Rev. Sci. Instrum.*, **18**, 108.
HODGINS, J. W., 1946, *Phys. Rev.*, **70**, 568.
HORN, F. H., and ZIEGLER, W. T., 1947, *J. Amer. Chem. Soc.*, **69**, 2762.
HULL, R. A., 1947, *Report of Cambridge Conference on Low Temperatures* (London : Physical Society), p. 72.
VAN ITTERBEEK, A., 1939, *Nature, Lond.*, **143**, 560.
VAN ITTERBEEK, A., VAN PAEMEL, O., and MARIENS, P., 1943, *Physica*, **10**, 90.
JOHNSON, A. F., 1947, *Canad. J. Res.*, **25**, 299.
JONES, G. O., LARSEN, A. H., and SIMON, F. E., 1948, *Research*, **1**, 420.
JONES, H., 1939, *Rep. Prog. Phys.*, **6**, 280.
KAPITZA, P., 1934, *Proc. Roy. Soc.* A, **147**, 189; 1941a, *J. Phys. U.S.S.R.*, **4**, 181; 1941 b, *Ibid.*, **5**, 59.
KEESOM, W. H., and DUYKAERTS, G., 1947, *Physica*, **13**, 153.
KEESOM, W. H., and KEESOM, P. H., 1941, *Physica*, **8**, 65.
KEESOM, W. H., SARIS, B. F., and MEYER, L., 1940, *Physica*, **7**, 877.
KEESOM, W. H. and WESTERMIJZE, W. K., 1941, *Physica*, **8**, 1044.
KISTEMAKER, J., 1947, *Physica*, **13**, 81.
KOPPE, H., 1947, *Ann. Phys., Lpz.*, **1**, 405; 1948, *Z. Naturforsch.*, **3a**, 1.
LANDAU, L., 1941, *J. Phys. U.S.S.R.*, **5**, 71; 1944, *Ibid.*, **8**, 1; 1947, *Ibid.*, **11**, 91.
LANDAU, L., and POMERANCHUK, I., 1948, *C.R. Acad. Sci. U.S.S.R.*, **59**, 669.
LANE, C. T., FAIRBANK, H. A., ALDRICH, L. T., and NIER, A. O., 1948, *Phys. Rev.*, **73**, 256.
LANE, C. T., FAIRBANK, H. A., and FAIRBANK, W. M., 1947, *Phys. Rev.*, **71**, 600.
LARSEN, A. H., SIMON, F. E., and SWENSON, C. A., 1948, *Rev. Sci. Instrum.*, **19**, 266.
v. LAUE, M., 1947 a, *Theorie der Supraleitung* (Berlin : Springer); 1947 b, *Report of Cambridge Conference on Low Temperatures* (London : Physical Society), p. 90; 1947 c, *Naturwissenschaften*, **34**, 186.
LAUMANN, E., and SHOENBERG, D., 1947, *Nature, Lond.*, **160**, 747.
DE LAUNAY, J., and DOLECEK, R., 1947, *Phys. Rev.*, **72**, 141.
LIFSHITZ, E., 1944, *J. Phys. U.S.S.R.*, **8**, 2.
LONDON, F., 1946, *Rev. Mod. Phys.*, **17**, 310; 1947, *Report of Cambridge Conference on Low Temperatures* (London : Physical Society), p. 1; 1948, *Phys. Rev.*, **74**, 562.
LONDON, F., and RICE, O. K., 1948, *Phys. Rev.*, **73**, 1188.
LONDON, H., 1939, *Proc. Roy. Soc.* A, **171**, 484; 1947 a, *Report of Cambridge Conference on Low Temperatures* (London : Physical Society), p. 109; 1947 b, *Ibid.*, p. 48.
MACDONALD, D. K. C., and MENDELSSOHN, K., 1948 a, *Nature, Lond.*, **161**, 972; 1948 b, *Ibid.*, **162**, 924.
MACDONALD, D. K. C., MENDELSSOHN, K., and BIRCH, A. J., 1947, *Phys. Rev.*, **71**, 563.
MELLINK, J. H., 1947, *Physica*, **13**, 180.
MENDELSSOHN, K., 1945, *Proc. Phys. Soc.*, **57**, 371; 1946, *Rep. Prog. Phys.*, **10**, 358 (London : Physical Society); 1947 a, *Report of Cambridge Conference on Low Temperatures* (London : Physical Society), p. 35; 1947 b, *Ibid.*, p. 41.
MENDELSSOHN, K., and PONTIUS, R. B., 1936, *Nature, Lond.*, **138**, 29.
MESHKOVSKY, A., and SHALNIKOV, A., 1947, *J. Phys. U.S.S.R.*, **11**, 1.
MEYER, L., and BAND, W., 1948, *Phys. Rev.*, **74**, 394.
MEYER, L., and MELLINK, J. H., 1947, *Physica*, **13**, 197.
OGG, R. A., 1946 a, *Phys. Rev.*, **69**, 243, 544; 1946 b, *Ibid.*, **70**, 93.
OSBORNE, D. V., 1948, *Nature, Lond.*, **162**, 213.
PELLAM, J. R., 1948, *Phys. Rev.*, **74**, 841.
PELLAM, J. R., and SQUIRE, C. F., 1947, *Phys. Rev.*, **72**, 1245.
PESHKOV, V., 1944, *J. Phys. N.K.R.*, **8**, 381; 1948, *J. exp. theor. Phys. U.S.S.R.*, **18**, 951.
PHILBERT, G., 1946, *J. Phys. Radium*, **7**, 243.
PICKARD, G. N., and SIMON, F. E., 1948, *Proc. Phys. Soc.*, **60**, 405.
PIPPARD, A. B., 1947, *Proc. Roy. Soc.* A, **191**, 370; 1948, *Nature, Lond.*, **162**, 68.

RANQUE, G., 1933, *J. Phys. Radium*, **4**, 112.
ROLLIN, B. V., and HATTON, J., 1948, *Phys. Rev.*, **74**, 508.
SALVETTI, C., 1941, *Ric. Sci.*, **12**, 894.
SCHALLAMACH, A., 1943, *J. Sci. Instrum.*, **20**, 195.
SCHIFF, L. I., 1941, *Phys. Rev.*, **59**, 751, 758.
SCHUBERT, G., 1946, *Z. Naturforsch.*, **1**, 111.
SHOENBERG, D., 1947, *Nature, Lond.*, **159**, 303.
SIENKO, M. J., and OGG, R. A., 1947, *Phys. Rev.*, **71**, 319.
SIMON, F., 1926, *Phys. Z.*, **27**, 790.
SLATER, J. C., 1948, *Proc. Phys. Soc.*, **61**, 577.
SOMMERFELD, A., 1942, *Ber. dtsch. chem. Ges.*, **75**, 1988.
STOUT, J. W., 1947, *Phys. Rev.*, **71**, 741.
SYDORIAK, SG., GRILLY, E. R., and HAMMEL, E. F., 1949, *Phys. Rev.*, in press.
TISZA, L., 1938, *Nature, Lond.*, **141**, 913; 1947 a, *Phys. Rev.*, **72**, 838; 1947 b, *Physics Today*, **1**, No. 4, 4.
TUMANOV, K. A., SHALNIKOV, A. I., and SHARVIN, J. V., 1947, *C.R. Acad. Sci. U.S.S.R.*, **56**, 35.
WEBBER, R. T., and REYNOLDS, J. M., 1948, *Phys. Rev.*, **73**, 640.
WEXLER, A., 1946, *Phys. Rev.*, **70**, 219.
YVON, J., 1948, *C.R. Acad. Sci., Paris*, **227**, 763.

SLOW NEUTRON ABSORPTION CROSS SECTIONS OF THE ELEMENTS

By M. ROSS and J. S. STORY
Theoretical Division, Atomic Energy Research Establishment, Harwell

CONTENTS

	PAGE
Explanation of table	291
Table of cross sections	293
References and notes	301

EXPLANATION OF TABLE

THE following list of published values for slow neutron absorption cross sections is not intended to be exhaustive, but contains the most recent and reliable declassified data that we are aware of on 1st July 1949.

The neutron absorption cross section σ_a is defined as the sum of the cross sections for the following nuclear reactions

$$(n, \gamma),\ (n, p),\ (n, d),\ (n, \alpha),\ (n, \text{fission});$$

more complex reactions are impossible or very improbable with thermal neutrons and so may be disregarded. In general, with thermal neutrons, reactions emitting charged heavy particles take place only in some of the lightest elements, [e.g. $^{10}B\ (n, \alpha)\ ^{7}Li$, $^{14}N\ (n, p)\ ^{14}C$]. With medium or heavy target nuclei, the potential barrier becomes too high and the compound nucleus is much more likely to lose its excitation energy by emitting gamma-rays.

The various results under each element are arranged in order of preference, in so far as it is possible to make any choice. The most satisfactory experimental methods for obtaining the absorption cross sections are:

(1) *Pile reactivity* ("danger coefficient"); cf. refs. (12), (48).

These methods are probably the most reliable. Some of the results quoted in Segrè's isotope chart [ref. (6)] and in the tables in refs. (29) (76) may have been obtained by pile reactivity experiments, but no bibliography is yet available for these three collections of data.

(2) *Velocity selectors and neutron spectrometers.*

These yield total neutron cross sections σ_{tot}. In many cases it is possible to separate the results into a constant term, presumed to be due to potential scattering, and a term varying with neutron energy (usually inversely proportional to the neutron velocity v in the thermal region), which is presumed to be due to absorption processes.

When $\sigma_a \ll \sigma_{sc}$ this separation is not always reliable owing to the possibility of molecular and crystal scattering effects contributing to the $1/v$ term in the thermal region, e.g. Hydrogen, ref. (56), Fluorine, ref. (42). It is probably from this cause that absorption cross sections obtained by this method are often rather higher than those obtained by other methods.

In some cases such transmission experiments yield at best an upper limit ($\sigma_{tot} > \sigma_a$) to the absorption cross section, e.g. Rhenium.

(3) *Activation studies*, cf. refs. (25), (37).

These methods are subject to many uncertainties discussed in ref. (25). Decay schemes for the activated isotopes are not always well known. Counting rates have to be corrected for beta energies and for gamma-ray counts, and may be too high owing to conversion electrons [e.g. Sodium, ref. (25)], or too low owing to unsuspected decays by orbital electron capture [e.g. Europium, ref. (25) (*y*)]. Neutron capture may lead in some cases to undiscovered long-lived isomeric states of the radioactive nucleus, or to isomeric (gamma ray) transitions [treated in error as beta-decays, e.g. Dysprosium, ref. (25)] to long-lived beta emitters. In many cases we find that cross sections measured by an activation method are rather lower than the results of the two previous methods. Nevertheless in cases for which $\sigma_a \ll \sigma_{sc}$ activation studies sometimes yield the most reliable data, since it is possible to separate the effects of contaminants by study of the half-lives of the resulting activities. Moreover, the method is valuable when only microscopic amounts of the material are available.

Of course for many elements having several stable isotopes activation methods are capable of yielding only a lower limit to the absorption cross section, e.g. Zirconium.

Effective neutron energy.

Of the various methods employed for measuring absorption cross sections [e.g. refs. (1), (23), (30), (32), (48)], some rely on comparison with a standard absorber such as Boron or Manganese: in such cases we have adjusted the experimental results quoted to the best value of the comparison cross section obtainable from this present collection of data (see footnotes). Many of the older results were obtained by the "Cadmium difference" technique which has since been shown to yield effective neutron energies rather above thermal, say $\sim$0·041 ev. [see refs. (16), (40)]. We have assumed that this is the case with the data of ref. (32), and, with elements whose slow neutron capture cross sections obey the $1/v$ law, have corrected the results to a neutron energy of 0·0253 ev. In other cases where the "Cadmium difference" technique has been employed [e.g. refs. (23), (30)] the same effect has been achieved (and probably more accurately) by correcting the value of the comparison cross section as mentioned above.

The word "thermal" in the third column has been used to imply a degree of doubt in the effective neutron energy (perhaps 0·0253 ev. $\pm$ 20%).

Probable errors.

This procedure should make the various results more strictly comparable, and we are able to obtain some idea of the probable errors. It is apparent from the table that in very few cases can we consider σ_a to be determined within a standard deviation better than 15%.

In general, when good agreement is obtained by methods (1) (2) and (3) it seems best to disregard earlier experiments. The earlier data are then, however, sometimes included (this applies especially to ref. (32)] for comparison so that we may gauge the probable error in cases for which more recent data are not available. In a few cases we have indicated the most likely value of σ_a by a rectangle but in general we avoid this selection.

Element	σ_a (barns)	Type of neutron	Method Remarks	References (see pp. 301–304)
1 H	$\boxed{0{\cdot}313 \pm 0{\cdot}013}$	2200 m/sec.	Diffusion in aqueous solutions	(1)
	$0{\cdot}313 \pm 0{\cdot}010$			(2)
	$0{\cdot}363 \pm 0{\cdot}012$	2200 m/sec.		(3)
	$0{\cdot}326 \pm 0{\cdot}014$	2200 m/sec.	Diffusion in aqueous solutions	(4)
	0·33	2200 m/sec.	Neutron lifetime in water	(5)
	$0{\cdot}31_8$	Thermal	Diffusion	(32)
1 ^{2}D	0·00032 (isotopic x-section)	Pile	Activation (?)	(6) (*I*)
D_2O	$(9{\cdot}2 \pm 2{\cdot}2) \times 10^{-4}$	2200 m/sec.	Diffusion length in D_2O	(7)
2 He	(atmospheric)			
	$(4{\cdot}6 \pm 0{\cdot}4) \times 10^{-3}$	Thermal column	Gridded ionization chamber	(77)
	$(6{\cdot}6 \pm 1{\cdot}0) \times 10^{-3}$	2200 m/sec.	Pulse analyser	(83)
	$(6{\cdot}6 \pm 1{\cdot}1) \times 10^{-3}$	2200 m/sec.	Transmission	(89)
	Due to ^{3}He (n, p) reaction.			
3 Li	$67{\cdot}3 \pm 2{\cdot}2$	2200 m/sec.	Diffusion in aqueous solutions	(8)
	$72{\cdot}3 \pm 1{\cdot}3$	2200 m/sec.	Vel. spectr.	(9)
	$59{\cdot}_2 \pm 9{\cdot}_7$	2200 m/sec.	Diffusion length in D_2O solution	(10)
	61	2200 m/sec.		(11)
	74	Thermal	Diffusion	(32)
	$\boxed{67{\cdot}_6 \pm 7{\cdot}_4}$	2200 m/sec.	Weighted average of above	
	Mainly due to the ^{6}Li (n, α) reaction.			
4 Be	$\boxed{0{\cdot}0085}$	2200 m/sec.	Pile reactivity	(12)
	0·010		Pile reactivity	(13)
	$<0{\cdot}021$	Thermal	Diffusion length in Be	(65)
5 B	698	2200 m/sec.	Vel. selector	(14)
	699	2200 m/sec.	Vel. selector	(14)
	702 ± 15	2200 m/sec.	In foil	(15)
	706 ± 16	2200 m/sec.	Vel. spectr.	(16)
	720	2200 m/sec.	Vel. spectr.	(17)
	733 ± 9	2200 m/sec.	Vel. spectr.	(18)
	743 ± 25	2200 m/sec.	Vel. selector	(19)
	$\boxed{710 \pm 21}$	2200 m/sec.	Weighted average of above	(*G*)
	Mainly due to the ^{10}B (n, α) reaction.			
6 C	$\boxed{0{\cdot}0045}$	Thermal		(6)
	0·0049	2200 m/sec.	Diffusion	(12)
	0·00592	2200 m/sec.	(AGX graphite)	(20)
	$0{\cdot}0060 \pm 0{\cdot}0010$	Thermal	Diffusion	(21)

σ_a(C) clearly decreases as the purity of the carbon is improved and would probably be even lower for spectroscopically pure material.

Element	σ_a (barns)	Type of neutron	Method Remarks	References (see pp. 301–304)
7 N	$\approx\sigma_{np}=1{\cdot}78\pm0{\cdot}07$	2200 m/sec.	Pulse analyser	(82)
	1·74	Thermal		(81) (6)
	$1{\cdot}25\pm0{\cdot}07$	2200 m/sec.	Diffusion in aqueous solutions	(22)
	$1{\cdot}45\pm0{\cdot}11$	2200 m/sec.	Neutron density in aqueous solutions	(23)
	1·34	Thermal	Diffusion	(32)
	Mainly due to the ^{14}N (n, p) reaction.			
8 O	$(2{\cdot}8\pm2{\cdot}2)\times10^{-4}$	Pile	D_2O–2D	(6) (*l*)
	$\gtrsim(1{\cdot}9\pm0{\cdot}5)\times10^{-4}$	Thermal column	^{17}O (n, α)	(24)
	0·0016	Thermal	See note (*b*)	(6)
9 F	$0{\cdot}0094\pm0{\cdot}0015$	Pile	Activation	(25) (6)
	$\lesssim0{\cdot}01$	~0·040 ev.		(26)
	0·01	~0·041 ev.	Activation	(27)
	0·064	Thermal	Diffusion	(32)
	0·065	Thermal		(6)
10 Ne	$<\sigma_{tot}=2{\cdot}79\pm0{\cdot}14$	~0·041 ev.		(28) (6)
	($\sigma_{sc}\approx2{\cdot}5$)			(29)
11 Na	0·5	Thermal		(6)
	0·41	Pile	Activation	(80)
	$0{\cdot}56\pm0{\cdot}05$	2200 m/sec.	Neutron density in graphite powder	(30)
	$\lesssim0{\cdot}63\pm0{\cdot}13$	Pile	Activation	(25) (*m*)
	0·48	2200 m/sec.	Activation	(31)
	0·48	Thermal	Diffusion	(32)
12 Mg	0·4	Thermal		(6)
	$0{\cdot}32\pm0{\cdot}03$	2200 m/sec.	Neutron density in graphite powder	(30)
	$0{\cdot}18\pm0{\cdot}05$	2200 m/sec.	Diffusion in aqueous solutions	(22)
	0·39	Thermal	Diffusion	(32)
13 Al	$0{\cdot}21\pm0{\cdot}04$	Thermal column	Activation	(25)
	$0{\cdot}02\pm0{\cdot}04$	2200 m/sec.	(Ra-Be) source in aqueous solutions	(33)
	$0{\cdot}22\pm0{\cdot}01$	2200 m/sec.	Neutron density in graphite powder	(30)
	0·29	2200 m/sec.	Activation	(31)
14 Si	0·25	Thermal		(6)
	<0·08	Thermal	Diffusion	(32)
	$\lesssim0{\cdot}01$	~0·04 ev.		(26)
15 P	0·31	Thermal		(6)
	$0{\cdot}23\pm0{\cdot}05$	Pile	Activation	(25)
	~0·27	2200 m/sec.	Activation	(31)
	0·29	Thermal	Diffusion	(32)

Element	σ_a (barns)	Type of neutron	Method Remarks	References (see pp. 301–304)
16 S	$0{\cdot}49 \pm 0{\cdot}05$	2200 m/sec.	Pile reactivity	(86)
	$0{\cdot}62 \pm 0{\cdot}06$	2200 m/sec.	Vel. spectr.	(42)
	$0{\cdot}37 \pm 0{\cdot}04$	2200 m/sec.	Diffusion in aqueous solutions	(22)
	$0{\cdot}41 \pm 0{\cdot}04$	2200 m/sec.	Neutron density in graphite powder	(30)
	0·78	Thermal	Diffusion	(32)
	0·53	Thermal		(6)
17 Cl	33	Thermal		(6)
	$32{\cdot}_3 \pm 1{\cdot}_4$	2200 m/sec.	Neutron density in aqueous solutions	(23)
	$34 \pm 2{\cdot}_6$	2200 m/sec.	Neutron density in graphite powder	(30)
	31	Thermal	Diffusion	(32)
18 A	0·77	2200 m/sec.	Vel. spectr.	(36)
	0·62	Thermal		(6)
	1·24	Thermal	Activation	(6)
19 K	2·2	Thermal		(6)
	$3{\cdot}8 \pm 0{\cdot}3$	2200 m/sec.	Neutron density in aqueous solutions	(23)
	~1·9	2200 m/sec.	Activation	(31)
	2·5	Thermal	Diffusion	(32)
20 Ca	$0{\cdot}44 \pm 0{\cdot}05$	2200 m/sec.	Neutron density in graphite powder	(30) (6)
	$0{\cdot}52 \pm 0{\cdot}06$	2200 m/sec.	Diffusion in aqueous solutions	(22)
	0·45	Thermal	Diffusion	(32) (*J*)
21 Sc	22 ± 4	Pile	Activation	(25) (*q*)
	>16	Thermal	Activation	(37)
22 Ti	$4{\cdot}5 \pm 0{\cdot}5$	2200 m/sec.	Pile reactivity	(86)
	$6{\cdot}2 \pm 0{\cdot}4$	2200 m/sec.	Neutron density in graphite powder	(30)
	4·8	Thermal	Diffusion	(32)
	5·2	0·035 ev. perhaps		(6) (*t*)
23 V	$4{\cdot}5 \pm 0{\cdot}9$	Thermal column	Activation	(25)
	5	Thermal		(29)
	9	Thermal		(6)
	$8{\cdot}1 \pm 0{\cdot}8$	2200 m/sec.	Neutron density in aqueous solutions	(23)
	~7·1	2200 m/sec.	Activation	(31)
24 Cr	$3{\cdot}0 \pm 0{\cdot}3$	Thermal	Pile oscillator	(91)
	3·4	2200 m/sec.	Vel. spectr.	(46)
	$3{\cdot}0 \pm 0{\cdot}1$	2200 m/sec.	Neutron density in graphite powder	(30)
	3	Thermal		(6)

Element	σ_a (barns)	Type of neutron	Method Remarks	References (see pp. 301–304)
25 Mn	$14{\cdot}1 \pm 0{\cdot}3$	2200 m/sec.	Vel. spectr.	(35)
	$10{\cdot}7 \pm 2{\cdot}4$	Thermal column	Activation	(25)
	$13{\cdot}1 \pm 0{\cdot}7$	2200 m/sec.	Diffusion in aqueous solutions	(38)
	$12{\cdot}7 \pm 0{\cdot}4$	2200 m/sec.	Neutron density in graphite powder	(30)
	$11{\cdot}1 \pm 0{\cdot}6$	2200 m/sec.	Neutron density in aqueous solutions	(23)
	14·2	Thermal	Diffusion	(32)
	$\boxed{12{\cdot}8 \pm 1{\cdot}6}$	2200 m/sec.	Weighted average of above	(6)
26 Fe	2·5	2200 m/sec.	Pile reactivity	(39)
	2·5	Thermal	Pile oscillator	(48)
	2·2	2200 m/sec.	Vel. spectr.	(40)
	2·4	2200 m/sec.	Ra-Be source in aqueous solutions	(33)
	$2{\cdot}43 \pm 0{\cdot}19$	2200 m/sec.	Neutron density in graphite powder	(30)
	2·14	Thermal		(6)
	2·0	Thermal	Diffusion	(32)
27 Co	$38{\cdot}_8$	2200 m/sec.	Vel. spectr.	(74)
	$40{\cdot}2 \pm 0{\cdot}9$	2200 m/sec.	Vel. spectr.	(41)
	$22{\cdot}4 \pm 4{\cdot}5$	Pile	Activation	(25)
	$39{\cdot}_2 \pm 3{\cdot}_5$	2200 m/sec.	Neutron density in aqueous solutions	(23)
	30	Thermal	Diffusion	(32)
	33	~0·035 ev. perhaps		(6) (*u*)
28 Ni	4·2	Thermal	Pile oscillator	(48)
	4·8	2200 m/sec.	Vel. spectr.	(42)
	$6{\cdot}6 \pm 0{\cdot}6$	2200 m/sec.	Neutron density in paraffin	(30)
	4·6	Thermal	Diffusion	(32)
	4·4	Thermal		(6)
	Mainly due to neutron capture in ^{58}Ni, see refs. (76) (74).			
29 Cu	$3{\cdot}6_3 \pm 0{\cdot}1_8$	Thermal	Pile oscillator	(91)
	3·4	2200 m/sec.	Vel. spectr.	(39)
	$2{\cdot}5_6 \pm 0{\cdot}5_1$	Thermal column	Activation	(25)
	$4{\cdot}0 \pm 0{\cdot}4$	2200 m/sec.	Neutron density in paraffin	(30) (6)
	$3{\cdot}0 \pm 0{\cdot}3$	Thermal	Diffusion	(32)
30 Zn	$1{\cdot}09 \pm 0{\cdot}08$	2200 m/sec.	Neutron density in paraffin	(30)
	0·9	Thermal		(76)
	1·25	Thermal		(6)
	1·5	Thermal	Diffusion	(32)
31 Ga	$2{\cdot}1_6 \pm 0{\cdot}4_{13}$	Pile	Activation	(25)

Element	σ_a (barns)	Type of neutron	Method Remarks	References (see pp. 301–304)
32 Ge	~0·6			(29)
	>0·25	Pile	Activation	(43)
	$<\sigma_{tot}\simeq 6$	2200 m/sec.	Vel. spectr.	(41)
	2·8	Thermal		(76)
33 As	5·6			(6)
	4·2±0·8	Thermal column	Activation	(25)
	7·7±0·4	2200 m/sec.	Neutron density in aqueous solutions	(23)
	4·2	2200 m/sec.	Activation	(31)
	4·4	Thermal	Diffusion	(32)
34 Se	16			(6)
	13·6±0·5	2200 m/sec.	Neutron density in aqueous solutions	(23)
	15	Thermal	Diffusion	(32)
35 Br	6·6±1·3	Thermal column	Activation	(25)
	8·4±0·8	2200 m/sec.	Neutron density in aqueous solutions	(23)
	4·0	Thermal	Diffusion	(32)
36 Kr	$>0\cdot10_4$	Thermal	Activation	(44)
	$<\sigma_{tot}=27$			(6)
	0·1			(29)
37 Rb	0·55±0·11	Pile	Activation	(25)
38 Sr	1·5±0·1	2200 m/sec.	Neutron density in graphite powder	(30) (6)
	1·9	Thermal	Diffusion	(32)
39 Y	1·24±0·25	Pile	Activation	(25)
	$1\cdot2_3$	Thermal	Activation	(37)
40 Zr	0·32±0·08	Thermal	Pile oscillator	(91)
	0·4	Thermal		(76)
	<0·7	2200 m/sec.	Pile reactivity	(86)
	>0·098±0·036	Pile	Activation	(25)
	<0·44	Thermal	Diffusion	(32)
	Other published values are unreliable owing to the probability of Hf contamination. Published values are :—			
	4·7±0·6	2200 m/sec.	Vel. spectr.	(45)
	>0·5	Thermal		(6)
41 Nb, Cb	1·44	2200 m/sec.	Pile reactivity	(12)
	1·27±0·16	2200 m/sec.	Pile reactivity	(86)
	0·63±0·25	2200 m/sec.	Vel. spectr.	(41)
	~1±0·4	Thermal column	Activation	(20)
42 Mo	2·4	Thermal	Pile oscillator	(48)
	2·91±0·30	2200 m/sec.	Pile reactivity	(86)
	2·7±0·1	2200 m/sec.	Neutron density in graphite powder	(30)
	3·9	Thermal		(6)
	~80% due to capture in ^{95}Mo, see ref. (48).			

Element	σ_a (barns)	Type of neutron	Method Remarks	References (see pp. 301–304)
43 Tc	Unstable			
44 Ru	2·5	Thermal		(76)
	$>0{\cdot}51\pm0{\cdot}01$	Pile	Activation	(25) (*x*)
	$<\sigma_{tot}=8{\cdot}8$ ($\sigma_{se}\approx6{\cdot}6$	2200 m/sec. 1 to 5 ev.	Vel. selector Vel. selector)	(46)
45 Rh	142 ± 13	2200 m/sec.	Vel. spectr.	(90)
	149 ± 15	Thermal column	Activation	(25)
	$<\sigma_{tot}\approx146$	2200 m/sec.	Crystal spectr.	(47)
	($\sigma_{se}=6\pm2$)			(90)
46 Pd	7	Thermal		(76)
	9	Thermal		(6)
47 Ag	$55{\cdot}6\pm2{\cdot}2$	Thermal	Pile oscillator (Separated isotopes)	(48)
	$56{\cdot}9\pm0{\cdot}6$	2200 m/sec.	Vel. spectr.	(9)
	$59{\cdot}6\pm2{\cdot}3$	Thermal	Pile oscillator (natural silver)	(48)
	71 ± 14	Thermal column	Activation	(25)
	58·5	Thermal		(6)
48 Cd	$\boxed{2410\pm110}$	2200 m/sec.	Vel. spectr.	(35) (*w*)
Supporting results, but with lower resolution are:—				
	2200	2200 m/sec.	Crystal spectr.	(49)
	2790	2200 m/sec.	Crystal spectr.	(50)
Mainly due to capture in ^{113}Cd, see ref. (52). The peak of the resonance is at $0{\cdot}176\pm0{\cdot}002$ ev. where $\sigma_{tot}=7200\pm200$ b., see ref. (35).				
49 In	$\lesssim\sigma_{tot}=191$ ($\sigma_{se}\approx4$)	2200 m/sec.	Vel. selector	(53) (29)
Supporting results are obtained from the data of refs. (9), (45), (47).				
	$190{\cdot}6\pm4{\cdot}3$	Thermal	Pile oscillator	(75)
	190 ± 19	Thermal column	Activation	(25)
	186	Thermal	Diffusion	(32)
The first and main resonance peak is at $1{\cdot}44\pm0{\cdot}02$ ev. where $\sigma_{tot}=26400\pm6500$ b., see refs. (53), (45), (9).				
50 Sn	$0{\cdot}63\pm0{\cdot}06$	2200 m/sec.	Neutron density in graphite powder	(30)
	0·51	Thermal	Diffusion	(32)
	0·55	Thermal		(76)
	0·69	Thermal		(6)
51 Sb	4·7	Thermal		(6)
	$4{\cdot}02\pm0{\cdot}13$	2200 m/sec.	Vel. spectr.	(9)
	$4{\cdot}9\pm1{\cdot}0$	Pile	Activation	(25)
	$5{\cdot}7\pm0{\cdot}4$	2200 m/sec.	Neutron density in graphite powder	(30)
	5·6	Thermal	Diffusion	(32)
52 Te	4	Thermal		(76)
	5	Thermal		(6)
	3·7	Thermal	Diffusion	(32)

Element	σ_a (barns)	Type of neutron	Method Remarks	References (see pp. 301–304)
53 I	$7{\cdot}04 \pm 0{\cdot}31$	2200 m/sec.	Vel. spectr.	(41)
	$6{\cdot}66 \pm 0{\cdot}25$	2200 m/sec.	Vel. selector	(54)
	$\boxed{6{\cdot}85 \pm 0{\cdot}34}$	2200 m/sec.	Average of above	
	$6{\cdot}25 \pm 1{\cdot}25$	Thermal column	Activation	(25)
	$8{\cdot}1 \pm 0{\cdot}5$	2200 m/sec.	Neutron density in aqueous solutions	(23)
	6·3	Thermal	Diffusion	(32)
54 Xe	$< \sigma_{tot} = 35$	Thermal		(6)
	$> 0{\cdot}09$	Thermal	Activation	(76)
55 Cs	41	Thermal		(6)
	$25{\cdot}6 \pm 5{\cdot}1$	Pile	Activation	(25)
56 Ba	1·25	Thermal		(6)
	$1{\cdot}5 \pm 0{\cdot}1$	2200 m/sec.	Neutron density in graphite powder	(30)
	1·2	Thermal	Diffusion	(32)
57 La	7	Pile	Activation	(80)
	$8{\cdot}4 \pm 1{\cdot}7$	Pile	Activation	(25)
	$> 8{\cdot}5$	Thermal	Activation	(37)
	11·5	Thermal	Diffusion	(32)
	12	Thermal		(6)
58 Ce	$< \sigma_{tot} = 29$	Thermal		(6)
	$> 0{\cdot}37_5 \pm 0{\cdot}06_5$	Thermal	Activation	(79)
59 Pr	$10{\cdot}1 \pm 2{\cdot}0$	Thermal column	Activation	(25)
	16	Thermal	Activation	(37)
60 Nd	92	2200 m/sec.	Crystal spectr.	(55)
	90	Thermal		(6)
	46	Thermal		(76)
	Mainly due to capture in ^{143}Nd, for which $\sigma_a \sim 730$ b., see ref. (87).			
	$^{147}61_{3\cdot7y} \sim 60$	Pile	Activation	(57)
	(This is the longest lived isotope of element 61 so far known; half-life = 3·7 years.)			
62 Sm	6490	2200 m/sec.	Crystal spectr.	(58)
	8000	Thermal		(6)
	Mainly due to capture in ^{149}Sm, see refs. (59), (61). The principal resonance peak is at 0·096 ev., where $\sigma_{tot} = 15500$ b., see ref. (58). See, however, also refs. (85) and (25).			
63 Eu	4700	2200 m/sec.	Vel. selector and crystal spectr.	(58)
	2600	Thermal		(88)
	3960	Thermal	Activation	(37) (*y*)
	$\lesssim 1870 \pm 300$	Pile	Activation	(25) (*y*)
	Principal resonance peaks are at 0·465 ev. and about −0·011 ev. at which $\sigma_{tot} > 5670$ b. and $\sigma_{tot} \approx 5570$ b. respectively, see ref. (58).			
64 Gd	47000	2200 m/sec.	Vel. selector	(60)
	48000	2200 m/sec.	Crystal spectr.	(58)
	Mainly due to capture in ^{155}Gd and ^{157}Gd, see ref. (61). There is a resonance peak at 0·028 ev. at which $\sigma_{tot} = 45000$ b., see refs. (60), (58).			

Element	σ_a (barns)	Type of Neutron	Method Remarks	References (see pp. 301–304)
65 Tb	>37	Thermal	Activation	(25)+(37)
	$<\sigma_{tot}=1000$	Thermal		(76)
	$<\sigma_{tot}=15$	Thermal		(6)
66 Dy	$<\sigma_{tot}=1170$	2200 m/sec.	Vel. selector	(60)
	$<\sigma_{tot}=1180$	2200 m/sec.	Crystal spectr., see also ref. (55)	(58)
	~ 940	Thermal	Diffusion	(32)
	850	($\sim$0·041 ev. perhaps)		(6)
	$>758\pm86$	Thermal column	Activation	(25) (*z*)
67 Ho	$59{\cdot}6\pm12$	Thermal column	Activation	(25)
	49	probably $\sim$0·4 ev.	Activation	(37)
	65	Thermal		(6)
68 Er	260	Thermal		(6)
	$<\sigma_{tot}=200$	Thermal		(76)
69 Tm	130	Thermal		(6)
	106 ± 11	Thermal column	Activation	(25)
	116	Thermal	Activation	(37)
70 Yb	$<\sigma_{tot}=65$	Thermal		(6)
	$>45\pm10$	Thermal	Activation	(64) (*A*)+(37)
71 Lu, Cp	195	Thermal		(6)
	$>125\pm22$	Thermal column	Activation	(25) (*B*)
	$>131\pm44$	Thermal	Activation	(64) (*C*) (*B*)
	$>138\pm20$	Thermal	Activation	(37) (*B*)
72 Hf	109 ± 16	2200 m/sec.		(86)
	$<\sigma_{tot}=130$	Thermal		(6)
	($\sigma_{se}=10$)			(29)
	$>3{\cdot}5\pm0{\cdot}7$	Pile	Activation	(25)
73 Ta	$18{\cdot}9\pm0{\cdot}6$	2200 m/sec.	Vel. spectr.	(45)
	$20{\cdot}6\pm4{\cdot}1$	Thermal column	Activation	(25) (29)
	35	Thermal		(6)
	15	Thermal	Diffusion	(32)
74 W	16	Thermal		(6)
	$17{\cdot}1\pm3{\cdot}1$	2200 m/sec.	Vel. spectr.	(45)
	$>10{\cdot}8\pm2{\cdot}2$	Thermal column	Activation	(25)
	26	Thermal	Diffusion	(32)
75 Re	$<\sigma_{tot}=97$	2200 m/sec.	Crystal spectr.	(66)
	85 ± 17	Thermal column	Activation	(25)
	130	Thermal		(6)
	132	Thermal	Diffussion	(32)
76 Os	$17{\cdot}0\pm0{\cdot}6$	2200 m/sec.	Vel. spectr.	(41)
	20	Thermal		(6)
	15	Thermal		(76)
77 Ir	402 ± 13	2200 m/sec. See also refs. (49) (58)	Vel. spectr.	(35)
	467 ± 93	Thermal column	Activation	(25)
	400	Thermal		(6)

Element	σ_a (barns)	Type of neutron	Method Remarks	References (see pp. 301–304)
78 Pt	10·8	Thermal		(6)
	6·5±0·4	2200 m/sec.	Vel. spectr.	(45)
	15	Thermal		(76)
79 Au	94·7±1·9	2200 m/sec.	Vel. spectr.	(67) (6)
	See also refs. (45), (9), (58), (60).			
	96·4±9·6	Thermal column	Activation	(25) (*D*)
	91·6	Thermal	Diffusion	(32)
80 Hg	377±18	2200 m/sec.	Vel. spectr.	(9)
	414	Thermal	Diffusion	(32)
	425	Thermal		(6)
	430	Thermal		(76)
	Mainly due to capture in ^{199}Hg; ref. (68) (*E*).			
81 Tl	2·9	Thermal		(6)
	3·8±1·3	2200 m/sec.	Vel. spectr.	(41)
	2·3±0·5	Pile	Activation	(25)
	2·8	Thermal	Diffusion	(32)
82 Pb	0·17	Thermal		(6)
	0·18	2200 m/sec.	Vel. spectr.	(69)
	0·23±0·06	2200 m/sec.	Ra-Be source in hydrogenous medium	(22)
	0·27±0·07	2200 m/sec.	Neutron density in paraffin	(30)
	0·38	Thermal	Diffusion	(32)
83 Bi	0·015±0·003	Pile	Activation	(25)
	0·016	Thermal		(6)
	$0{\cdot}013_5$±0·001$_7$	Thermal	Activation	(70)
90 ^{232}Th	$\sigma_{n\gamma}$=6	0·041 ev. probably		(6) (29) (*F*)
	=9·0	Thermal	Diffusion	(32)
92 U	$\sigma_c \sim 2$			(29)
	$\sigma_{fiss} \approx 3{\cdot}6$	Thermal	"pre 1941 literature"	(29)
93 ^{237}Np	$\sigma_c \sim 200$	Thermal		(72)
	$\sigma_{fiss} \sim 0{\cdot}02$	Thermal		(73)

ACKNOWLEDGMENTS

We are indebted to Sir John Cockcroft, Director, A.E.R.E., for permission to publish this paper, and to Messrs. F. C. W. Colmer and E. R. Wiblin for some hitherto unpublished results.

REFERENCES AND NOTES

Bibliographical references are numbered, and footnotes run in the sequence (*a*) . . . . (*z*), (*A*), (*B*), ...

(1) $\sigma_a(B)/\sigma_a(H)=2270\pm63$, (*a*); WHITEHOUSE, W. J., and GRAHAM, G. A. R. (1945), *Can. J. Res.*, **A 25**, 261 (1947).

(2) $\sigma_a(B)/\sigma_a(H)=2270$, (*a*); WALKER, R. L., *M.D.D.C.* 414.

(3) $\sigma(B)/\sigma_a(H)=1955\pm24$, (*a*); SCHULZ, L. G., and GOLDHABER, M., *Phys. Rev.*, **67,** 202 (1945).

(4) $\sigma_a(B)/\sigma_a(H)=2180\pm3\%$, (*a*); REEL, H. L., and GROSJEAN, C., *C.R. Acad. Sci., Paris*, **226,** 1598 (1948).

No correction was made for perturbation of the neutron distribution by the detectors.

(5) MANLEY, J. H., HAWORTH, L. J., and LUEBKE, E. A., *Phys. Rev.*, **61,** 152 (1942).

(6) Quoted in " Isotope Chart ", SEGRÈ, E., revised April 1948; *A.E.C.D.* 2111.

(7) (*b*), SARGENT, B. W., BOOKER, D. V., CAVANAGH, P. E., HEREWARD, H. G., and NIEMI, N. J. (1945), *Can. J. Res.*, **A 25,** 134 (1947).

(8) $\sigma_a(Li)/\sigma_a(B)=0{\cdot}0948\pm0{\cdot}0013$, (*a*); FENNING, F. W., GRAHAM, G. A. R., and SELIGMAN, H. (1943), *Can, J. Res.*, **A 25,** 73 (1947).

(9) (*c*), RAINWATER, J., and HAVENS, W. W., JR., *Phys. Rev.*, **70,** 154 (1946).

(10) HEREWARD, H. G., LAURENCE, G. C., MUNN, A. M., PANETH, H. R., and SARGENT, B. W. (1943) : *Can. J. Res.*, **A 25,** 26 (1947).

(11) (*d*), HARRINGTON, E. L., and STEWART, J. L., *Can. J. Res.*, **19,** 33 (1941).

(12) ANDERSON, H. L., FERMI, E., WATTENBERG, A., WEIL, G. L., and ZINN, W. H., *Phys. Rev.*, **72,** 16 (1947).

(13) HUGHES, D. J., EGGLER, C., and HUDDLESTON, C. M., *Phys. Rev.*, **71,** 269 (1947).

(14) (*c*), (*e*), (*f*), FERMI, E., MARSHALL, J., and MARSHALL, L., *Phys. Rev.*, **72,** 193 (1947).

(15) (*c*), (*e*), MARSHALL, J., *Phys. Rev.*, **70,** 107 (1947).

(16) (*e*), BACHER, R. F., BAKER, C. P., and MCDANIEL, B. D., *Phys. Rev.*, **69,** 443 (1946).

(17) (*c*), (*e*), SUTTON, R. B., MCDANIEL, B. D., ANDERSON, E. E., and LAVATELLI, L. S., *Phys. Rev.*, **71,** 272 (1947).

This result is for B in a BF_3 absorber.

(18) (*c*), (*e*), (*f*), (*g*), (*h*), *Ibid.* This is an average over the results for four different absorbers whose boron contents were somewhat in doubt.

(19) (*e*), (*f*), (*i*), (*j*), RAINWATER, J., and HAVENS, W. W., JR., *Phys. Rev.*, **70,** 136 (1946).

(20) Quoted by BERNSTEIN, S., *Phys. Rev.*, **73,** 956 (1948).

(21) (*g*), BOTHE, W., and JENSEN, P., *Z. Phys.*, **122,** 749 (1944).

(22) (*k*), FENNING, F. W., and SELIGMAN, H., *B.D.D.A.* 78 (1942).

(23) (*a*), LAPOINTE, C., and RASETTI, F., *Phys. Rev.*, **58,** 554 (1940).

(24) MAY, A. N., and HINCKS, E. P., *Can. J. Res.* **A 25,** 77 (1947).

(25) SEREN, L., FRIEDLANDER, H. N., and TURKEL, S. H., *Phys. Rev.*, **72,** 888 (1947).

(26) MUEHLHAUSE, C. O., and GOLDHABER, M., *Phys. Rev.*, **70,** 85 (1946).

(27) (*l*), MANLEY, J. H., HAWORTH, L. J., and LUEBKE, E. A., *Phys. Rev.*, **59,** 109 (1941).

(28) CARROLL, H., *Phys. Rev.*, **60,** 702 (1941).

(29) Quoted in : *The Science and Engineering of Nuclear Power* (Addison–Wesley Press, 1947), pp. 291–293.

(30) (*a*), COLTMAN, J. W., and GOLDHABER, M., *Phys. Rev.*, **69,** 411 (1946).

(31) (*n*), RASETTI, F., *Phys. Rev.*, **58,** 869 (1940).

(32) (*o*), VOLZ, H., *Z. Phys.*, **121,** 201 (1943).

(33) (*p*), DEMERS, P., *Can. J. Res.* **A 24,** 117 (1946).

(34) SEREN, L., MOYER, W. E., and STURM, W., *Phys. Rev.*, **70,** 561 (1946).

(35) HAVENS, W. W. JR., DUNNING, J. R., RAINWATER, L. J., and WU, C. S., *Phys. Rev.*, **71,** 65 (1947).

(36) MELKONIAN, E., RAINWATER, L. J., HAVENS, W. W. JR., and DUNNING, J. R., *Phys. Rev.*, **73,** 1399 (1948).

(37) (*r*), BOTHE, W., *Z. Naturforsch.*, **1,** 179 (1946).

(38) (*a*), (*s*), $\sigma_a(Mn)/\sigma_a(H)=42\pm1{\cdot}5$; BALLINI, R., BERTHELOT, A., and SMEETS, C., *C.R. Acad. Sci., Paris*, **225,** 328 (1947).

(39) HUGHES, D. J., WALLACE, J. R., and HOLTZMAN, R. H., *Phys. Rev.*, **73,** 1277 (1948).

(40) HAVENS, W. W., JR., RAINWATER, L. J., WU, C. S., and DUNNING, J. R., *Phys. Rev.*, **73,** 963 (1948).

(41) HAVENS, W. W., JR., RAINWATER, L. J., and WU, C. S., *Phys. Rev.*, **71,** 174 (1947).

(42) RAINWATER, L. J., HAVENS, W. W. JR., DUNNING, J. R., and WU, C. S., *Phys. Rev.*, **73,** 733 (1948).

(43) A correction of results of ref. (25) by ARNOLD, J. R., and SUGARMAN, N., *J. Chem. Phys.*, **15,** 703 (1947).

(44) HOAGLAND, E. J., and SUGARMAN, N., *M.D.D.C.* 614 L.

(45) Havens, W. W. Jr., Wu, C. S., Rainwater, L. J., and Meaker, C. L., *Phys. Rev.*, **71,** 165 (1947).
(46) Havens, W. W., Jr., and Rainwater, L. J. (1948), *A.E.C.D.* 2418.
(47) Borst, L. B., Ulrich, A. J., Osborne, C. L., and Hasbrouck, B., *Phys. Rev.*, **70,** 557 L (1946), see ref. (29).
(48) (*v*), Pomerance, M., and Hoover, J. I., *M.D.D.C.* 1650 (1948); see also *Phys. Rev.*, **73,** 1265 A (1948).
(49) (*w*), Sawyer, R. B., Wollan, E. O., Bernstein, S., and Peterson, K. C., *Phys. Rev.*, **72,** 109 (1947).
(50) (*w*), Zinn, W. H., *Phys. Rev.*, **71,** 752 (1947).
(51) Beeman, W. W., *Phys. Rev.*, **72,** 986 (1947).
(52) Moyer, B. J., Peters, B., and Schmidt, F. H., *Phys. Rev.*, **69,** 666 (1946); Dempster, A. J., *Phys. Rev.*, **71,** 829 (1947).
(53) McDaniel, B. D., *Phys. Rev.*, **70,** 832 (1946).
(54) Jones, W. B. Jr., *Phys. Rev.*, **72,** 362 (1947).
(55) Sturm, W. J., and Arnold, G. P., *Phys. Rev.*, **71,** 556 L (1947).
(56) Jones, W. B., Jr., *Phys. Rev.*, **74,** 364 (1948).
(57) Parker, G. W., Lantz, P. M., Inghram, M. G., Hess, D. C., Jr., and Hayden, R. J., *Phys. Rev.*, **72,** 85 (1947).
(58) Sturm, W. J., *Phys. Rev.*, **71,** 757 (1947).
(59) Dempster, A. J., *Phys. Rev.*, **74,** 505 (1948).
(60) Brill, T., and Lichtenberger, H. V., *Phys. Rev.*, **72,** 585 (1947).
(61) Lapp, R. E., Van Horn, J. R., and Dempster, A. J., *Phys. Rev.*, **71,** 745 (1947).
(62) O'Neal, R. D., and Goldhaber, M., *Phys. Rev.*, **59,** 102 L (1941).
(63) Inghram, M. G., Shaw, A. E., Hess, D. C., Jr., and Hayden, R. J., *Phys. Rev.*, **72,** 515 (1948).
(64) Atterling, H., Bohr, E., and Sigurgeirsson, Th., *Ark. Mat. Astr. Fys.*, **32 A,** No. 2 (1945).
(65) (*K*), Fünfner, E., and Bothe, W., *Z. Phys.*, **122,** 769 (1944).
(66) Sturm, W. J., and Arnold, G. P., *M.D.D.C.* 872 (declassified 1947).
(67) McDaniel, B. D., Sutton, R. B., Lavatelli, L. S., and Anderson, E. E., *Phys. Rev.*, **72,** 729 (1947).
(68) Inghram, M. G., Hayden, R. J., and Hess, D. C., Jr., *Phys. Rev.*, **71,** 561 (1947).
(69) Havens, W. W., Jr., Rabi, I. I., and Rainwater, L. J., *Phys. Rev.*, **72,** 634 (1947).
(70) (*n*), Maurer, W., *Naturwissenschaften*, **32,** 295 (1944).
(71) Meitner, L., *Nature, Lond.*, **145,** 422 (1940).
(72) Seaborg, G. T., *Science*, **104,** 386 (1946).
(73) Ghiorso, A., and Magnusson, L. B., *Science*, **104,** 386 (1946).
(74) Bernstein, S., Dial, J. B., Stephenson, T. E., Shapiro, M. M., and Stanford, C. P. (1948), *A.E.C.D.* 2404.
(75) (*v*), Hoover, J. I., Jordan, W. H., Moak, C. D., Pardue, L., Pomerance, H., Strong, J. D., and Wollan, E. O., *Phys. Rev.*, **74,** 864 (1948).
(76) See Way, K., and Haines, G. (1948), *A.E.C.D.* 2138.
(77) (*H*), Batchelor, R., Eppstein, J. S., Flowers, B. H., and Whittaker, A., *Nature, Lond.*, **163,** 211 (1949).
(78) Thode, H. G., Macnamara, J., Lossing, F. P., and Collins, C. B., *J. Amer. Chem. Soc.*, **70,** 3008 (1948).
(79) Katcoff, S., Leary, J. A., Walsh, K. A., Elmer, R. A., Goldsmith, S. S., Hall, L. D., Newbury, E. G., Povelites, J. J., and Waddell, J. S., *A.E.C.D.* 2174 (1948).
(80) Wattenberg, A., and Thomas, G.; see *Phys. Rev.*, **73,** 547 (1948).
(81) (*a*) Lichtenberger, H., Fowler, H., and Wattenberg, A. (1943), see *A.E.C.D.* 2190.
(82) (*a*) Coon, J. H., and Nobles, R. A., *Phys. Rev.*, **74,** 1358 (1949).
(83) (*H*), Coon, J. H., *Phys. Rev.*, **75,** 1355 (1949).
(84) Jenks, G. H., Ghormley, J. A., and Sweeton, F. H., *Phys. Rev.*, **75,** 701 L (1949).
(85) Sunyar, A. W., and Goldhaber, M., *Bull. Am. Phys. Soc.*, **24,** iv. 37 (1949).
(86) (*a*) Colmer, F. C. W., and Wiblin, E. R., 1948-1949 (private communication).
(87) Hess, D. C., and Inghram, M. G. (1948), *A.E.C.D.* 2470.
(88) Hayden, R. J., Reynolds, J. H., and Inghram, M. G., *Phys. Rev.*, **75,** 1500 (1949).

(89) (*a*), (*c*), (*e*), (*L*), KING, L. D. P. and GOLDSTEIN, L., *Phys. Rev.*, **75,** 1366 (1949).
(90) MEIJER, R. R., *Phys. Rev.*, **75,** 773 (1949).
(91) (*v*) POMERANCE, H. S., *A.E.C.D.* 2502 (1949).

B.D.D.A. = British declassified document (Atomic).
M.D.D.C. = Manhattan District declassification code.
A.E.C.D. = Atomic Energy Commission declassified documents.

(*a*) We assume $\sigma_a(B) = 710 \pm 21$ b., a weighted average from the results of refs. (14) to (19), (*G*).
(*b*) This value for $\sigma_a(D_2O)$ is less than the values given in ref. (6) for $\sigma_a(O)$ and also for $2\sigma_a(D)$. It is improbable that the heavy water used in this experiment was compounded principally with the heavy oxygen isotope, ^{18}O *, (since D_2O is generally prepared by electrolysis) so we must conclude that the values for $\sigma_a(D)$ and $\sigma_a(O)$ given in ref. (6) are both too large; see also note (*I*).
* $\sigma_a(^{18}O) = (2{\cdot}2 \pm 0{\cdot}44) \times 10^{-4}$ b., see ref. (34).
$\sigma_a(^{17}O) = 0{\cdot}49 \pm 0{\cdot}13$ b., ref. (24).
If the D_2O be enriched in ^{18}O it is likely to be enriched also (to a lesser degree) in ^{17}O, and so the neutron absorption in the oxygen content might well be increased.
(*c*) We assume $\sigma_{tot}(F) \simeq (3{\cdot}57 \pm 0{\cdot}22) + 0{\cdot}059E^{-\frac{1}{2}}$ b. (E in ev.).
(*d*) We assume $\sigma_a(Cd) = 2410 \pm 110$ b. at 2200 m/sec., see ref. (35) (*w*).
(*e*) We assume $\sigma_{sc}(B) = 3{\cdot}2 \pm 1{\cdot}2$ b., see refs. (14), (17), (16).
(*f*) We assume $\sigma_{tot}(O) = 3{\cdot}68$ b. for energies $\lesssim 2$ ev., refs. (36), (56); $\simeq 4{\cdot}1 \pm 0{\cdot}2$ b. at thermal energies.
(*g*) We assume $\sigma_{tot}(C) = 4{\cdot}7$ b. for energies $\gtrsim 1$ ev., refs. (56), (86); $\simeq 4{\cdot}8 \pm 0{\cdot}2$ b. at thermal energies.
(*h*) We assume $\sigma_{tot}(Pb) = 11{\cdot}2 \pm 0{\cdot}2$ b. for energies $\gtrsim 1$ ev.
(*i*) We assume $\sigma_{tot}(Al) = 1{\cdot}8 \pm 0{\cdot}1$ b. for thermal energies.
(*j*) Assumes $\sigma_{tot}(Si + O_2) \approx 10{\cdot}7$ b. for thermal energies; see *Phys. Rev.*, **70,** 136 (1946).
(*k*) We assume $\sigma_a(H) = 0{\cdot}313 \pm 0{\cdot}013$ b., see ref. (1).
(*l*) Assumes $\sigma_a(Al) = 0{\cdot}2$ b.
(*m*) σ_a may be too high owing to conversion electrons, see ref. (25).
(*n*) We assume $\sigma_a(Mn) = 12{\cdot}8 \pm 1{\cdot}6$ b.
(*o*) Effective neutron energy in these experiments is believed to be $\sim 0{\cdot}041$ ev. With elements for which $\sigma_a.v = \text{const.}$ in the thermal region we have made a correction to obtain the absorption cross sections at 0·0253 ev. (2200 m/sec.).
(*p*) We assume $\sigma_a(HO_4^-) = 0{\cdot}313 \pm 0{\cdot}013$ b., see ref. (1).
(*q*) Assumes 5% of decays are by K-capture, but σ may be too high owing to conversion electrons, see ref. (25).
(*r*) We assume $\sigma_{ac}(HO)_{27\ hr} = 59{\cdot}6 \pm 12$ b., see ref. (25).
(*s*) We assume $\sigma_a(B)/\sigma_a(H) = 2270 \pm 63$, see refs. (1), (2).
(*t*) This value in ref. (6) is possibly the result due to ref. (30) before the correction (*a*) is applied.
(*u*) This value in ref. (6) is possibly the result due to ref. (23) before the correction (*a*) is applied.
(*v*) We assume that $\sigma_a(Au) = 94{\cdot}7 \pm 1{\cdot}9$ b., see ref. (67).
(*w*) After subtraction of a potential scattering term $\sigma_{sc} = 5{\cdot}3 \pm 0{\cdot}7$ b., see ref. (35), and a resonance scattering term $\sigma_{res.\ sc} = 7{\cdot}2 \pm 1{\cdot}7$ b. at 2200 m/sec., see ref. (51).
(*x*) The 37 hr. decay observed in ref. (25) is presumably that of ^{106}Rh the daughter of 4·4 hr. ^{106}Ru, so the cross section is just that for ^{104}Ru again.
(*y*) We here assume 0·22 and 2·25 K-captures per beta decay for the isotopes $^{152}Eu_{9{\cdot}2\ hr.}$ and $^{153,154}Eu_{5{\cdot}4\ yr.}$ respectively, see ref. (88). The result is probably too large since some of the K x-rays will have been counted.
(*z*) This result is unreliable because ^{165}Dy (1·25 m.) was assumed to decay by beta emission. Later work, ref. (63), indicates that this isotope undergoes isomeric transition to the 2·3 hr. decay.
(*A*) We assume $\sigma_{ac}(^{176}Lu)_{3{\cdot}7\ hr.} = 28{\cdot}4 \pm 15$b., $\sigma_a(^{176}Lu) = 99 \pm 32$ b., per natural atom, a weighted average from refs. (25), (37) and (64).
(*B*) We assume ^{177}Lu (6·8 d.) decays by K-capture with 0·2 K-transitions per beta decay, see ref. (37).
(*C*) We assume $\sigma_{ac}(Dy)$ 2·3 hr. $= 700 \pm 110$ b., see refs. (25) and (64) (*n*), and $\sigma_{ac}(Lu)_{3{\cdot}7\ hr.} = 28{\cdot}4 \pm 15$ b., a weighted average from refs. (25), (37) and (64).
(*D*) A 14% correction for conversion electrons was made, see ref. (25).
(*E*) ^{198}Hg and ^{199}Hg were found to be the large absorbers, and the absolute values of their cross sections assigned to total 430 b. per natural atom, see ref. (76).
(*F*) Probably due to ref. (71) for neutrons of energy $\sim 0{\cdot}04$ ev.
(*G*) The observed variation in the abundance ratio $^{10}B/^{11}B$, ref. (78), makes a trivial contribution in the standard deviation computed here.
(*H*) We assume $\sigma(N) = 1{\cdot}78 \pm 0{\cdot}07\%$ b. for the (n, p) reaction at thermal energy, see ref. (82).
(*I*) In view of note (*b*) we have assumed that the value $\sigma_a(D) = 0{\cdot}65$ millibarns quoted in ref. (6) was obtained by an activation method; it was associated in earlier editions of (6) with a half-life 25 yrs. of the ^3H product nucleus; we here assume 12·46 yrs, see ref. (84).
(*J*) We assume $\sigma_a(F) = 0{\cdot}0094$ b., see ref. (25).
(*K*) We assume $\sigma_{sc}(Be) = 6{\cdot}1$ b , see ref. (76).
(*L*) We assume the abundance of ^3He in atmospheric He is $(1{\cdot}3_0 \pm 0{\cdot}2_0) \times 10^{-4}\%$, see ref. (83).

MOLECULAR DISTRIBUTION AND EQUATION OF STATE OF GASES

By J. de BOER
Institute for Theoretical Physics, University of Amsterdam

CONTENTS

Chapter I: Fundamental Problems

Page

§ 1. Introduction 306
§ 2. The kinetic method 307
(i) The kinetic velocity-distribution function 307
(ii) Boltzmann's equation 308
(iii) H-theorem 309
§ 3. Method of statistical mechanics 309
(i) Probability density of ensemble in γ-space 309
(ii) Change with time and equilibrium ensembles 311
(iii) Generic and specific phase. The μ-space 313
(iv) The probability density matrix in quantum statistical mechanics .. 314
(v) Time dependence and equilibrium ensembles in quantum statistical mechanics 317
§ 4. The problem of the justification of the method of statistical mechanics .. 318
(i) Equilibrium 318
(ii) Non-equilibrium 319

Chapter II: Molecular Distribution and Equation of State in Equilibrium

§ 5. Classical theory 325
(i) Canonical ensemble distribution functions 325
(ii) Caloric and thermal equation of state 326
(iii) Statistical thermodynamics with the canonical ensemble 327
§ 6. Quantum theory 329
(i) Canonical ensemble 329
(ii) Caloric and thermal equation of state 330
(iii) Correspondence between classical and quantum theory at high temperatures 332
§ 7. Series expansion in powers of the density, for distribution functions and equation of state 335
(i) Introduction 335
(ii) Development of equation of state in powers of z 336
(iii) Development of distribution functions $n_h(\mathbf{r}^h)$ in powers of z 338
(iv) Development of pV, U and $n_2(r_{12})$ in powers of the density 340
§ 8. Calculation of the equation of state 341
(i) Law of corresponding states 341
(ii) The experimental data 343
(iii) Classical theory 344
(iv) Quantum theory of the ideal gas 347
(v) Quantum theory of real gases at high temperatures 349
(vi) Quantum theory of real gases at very low temperatures 351
§ 9. Calculation of the molecular pair distribution function 354
(i) Series development into powers of the density 354
(ii) Distribution function and potential of average force 358
(iii) Integral equation for $n_2(r_{12})$ 362
§ 10. Fluctuations and scattering of light and x rays, in terms of the distribution functions 364
(i) Fluctuations 364
(ii) Scattering of electromagnetic waves 366
(iii) Molecular pair distribution function from x-ray analysis 368

§1. INTRODUCTION

THE properties of gases and liquids and their theoretical treatment by kinetic gas theory or the more powerful method of statistical mechanics, have acquired new interests in the last two decades, both because of new and more accurate experimental methods and because of renewed efforts to interpret the results of experiments in terms of intermolecular forces, as was first done by Lennard-Jones. Moreover, the development of quantum mechanics gave a new impetus to the study, both theoretical and experimental, of the influence of the wave character of the molecules on the behaviour of gases and liquids, to the reformulation of the laws of statistical mechanics and to the foundation of statistical mechanics in correspondence with the quantum mechanical formalism. Much of this has been described already in well-known textbooks as for instance those of Fowler (1936), Tolman (1938), Fowler and Guggenheim (1940), Mayer and Mayer (1946) on statistical mechanics, and the monograph of Chapman and Cowling (1939) on the kinetic theory of non-uniform gases.

Recently in a number of papers attention has been drawn again to the foundations of the kinetic gas theory and the method of statistical mechanics, especially in its application to systems which are not in a state of equilibrium. Therefore in §§ 2 and 3 a short discussion is given of these two methods both in classical and in quantum theory, while in § 4 some introductory remarks are made, to show the difficulties which arise in application to non-equilibrium processes, and to indicate the direction in which scientists expect the solution to be found.

This first chapter which is concerned mostly with the fundamentals of the kinetic gas theory and of the method of statistical mechanics, is followed by a second part, which is of a more practical nature. In order not to lengthen the treatment too much it was necessary to limit the treatment of the applications in two respects:

1. Only the application to the calculation of the distribution-function and the caloric and thermal equation of state of *gases* are discussed. This makes it impossible to discuss also the liquid state, as for example the very promising treatment of the properties of liquids by Lennard-Jones and Devonshire (1937, 1938), that of the hole-theory of liquids by Fürth (1941) and by Hirschfelder and collaborators (1937) and the calculation of molecular pair distribution functions of liquids starting from the crystalline state as given by Prins (1937), Bernal (1937), Wall (1940), Rushbrooke and Coulson (1939) and by Corner and Lennard-Jones (1941).

2. Only the properties of *systems in equilibrium* are discussed. So all applications to the stationary, but non-equilibrium, states which are considered when a temperature or velocity gradient is present are omitted. The promising contributions in this direction of Born and Green (1947 b), Kirkwood (1946, 1947), Eisenschitz (1949) and others are thus passed over in silence.

The general expressions, obtained with the aid of the canonical ensemble method of Gibbs for the distribution functions and the caloric and thermal equation of state, are given in §§ 5 and 6 for classical and quantum theory respectively. Here it is also shown that the expressions obtained are in agreement with the expressions obtained by applying the partition function method of statistical thermodynamics, both in the classical case and in the case when quantum theory has to be applied. In § 7 the caloric and thermal equation of state are developed into power series of powers of the density, using the general Ursell-Uhlenbeck procedure which is valid both in classical and quantum theory. A generalization

of this procedure is given, which allows one to obtain in a similar way the corresponding development of the distribution functions in powers of the density.

The next two sections, §§ 8 and 9, contain explicit calculations of the equation of state and of the distribution function when an intermolecular potential field of the Lennard-Jones type is introduced. Special attention is also paid to the influence of quantum theory on the equation of state and the distribution functions. In § 9 the connection between the molecular distribution functions and the potential of the average force acting on the molecules is discussed; this leads to an integral equation for the distribution function, which was derived in a different way by Born and Green (1946). Finally, in the last section, § 10, a survey is given which relates the molecular pair distribution function to the density fluctuations and to the intensity of light- or X-rays scattered by gases or liquids.

I. FUNDAMENTAL PROBLEMS

§ 2. THE KINETIC METHOD

(i) *The Kinetic Velocity-distribution Function*

For simplicity it is assumed that the gaseous or liquid system consists of molecules without internal (rotational or vibrational) degrees of freedom, which exert central intermolecular forces on each other. The instantaneous situation in this system of N molecules is specified by giving the number of molecules with (*a*) centre of gravity given by the position vector $\mathbf{r}_1$ inside the small but finite element space of $\delta\mathbf{r}_1 \equiv \delta x_1 \delta y_1 \delta z_1$, and with (*b*) linear momentum specified by the momentum vector $\mathbf{p}_1$ inside the small but finite element in momentum space $\delta\mathbf{p}_1 \equiv \delta p_{x1} \delta p_{y1} \delta p_{z1}$. The configuration and momentum spaces can be combined in *one* 6-dimensional μ-space, divided into small but finite elements $\delta\mathbf{r}_1, \delta\mathbf{p}_1$.

The element of volume $\delta\mathbf{r}_1$ must satisfy the following conditions: (i) it must be so large, that it contains a large number of molecules; (ii) it must be so small, that variations of macroscopic quantities, as for instance the density, the temperature or the hydrodynamic velocity, inside this volume element can be neglected (e.g. $\delta T/T \ll 1$). Because of the molecular motion and intermolecular collisions the number of molecules in $\delta\mathbf{r}_1 \delta\mathbf{p}_1$ changes with time. So the phrase "number of molecules at time t in $\delta\mathbf{r}_1 \delta\mathbf{p}_1$, surrounding $\mathbf{r}_1$, $\mathbf{p}_1$" has the meaning of an average number, averaged over a small but finite period of time δt, at t, which satisfies the condition (i) that it be large compared with the average time taken to cross the element $\delta\mathbf{r}_1$, and the condition (ii) that, on the other hand, it be small compared with the scale of time variation of macroscopic properties (cf. Chapman and Cowling 1939, § 2.2).

This time average of the number of molecules in the small but finite element $\delta\mathbf{r}_1$ surrounding $\mathbf{r}_1$ in space, and $\delta\mathbf{p}_1$ surrounding $\mathbf{p}_1$ in momentum space is denoted by $f(\mathbf{r}_1, \mathbf{p}_1, t)\delta\mathbf{r}_1\delta\mathbf{p}_1$. The function $f(\mathbf{r}, \mathbf{p}, t)$ is called the distribution function and it will be distinguished if necessary from other distribution functions in this paper by the adjective "*kinetic*" *distribution function*.

The distribution function satisfies the condition $\Sigma_i f(\mathbf{r}_i, \mathbf{p}_i, t)\delta\mathbf{r}_i\delta\mathbf{p}_i = N$, where the summation is carried out over all volume elements $\delta\mathbf{r}_i\delta\mathbf{p}_i$ of the μ-space. When the distribution function can be considered as a continuous function of $\mathbf{r}$ and $\mathbf{p}$, as is always assumed in kinetic gas theory, the summation over all elements

$\delta\mathbf{r}_i\delta\mathbf{p}_i$ can be replaced by an integration over $\mathbf{r}$ and $\mathbf{p}$, giving: $\iint f(\mathbf{r},\mathbf{p},t)\,d\mathbf{r}\,d\mathbf{p}=N$. When $f(\mathbf{r},\mathbf{p})$ does not depend on $\mathbf{r}$, $f(\mathbf{r},\mathbf{p})\,d\mathbf{p}=f(\mathbf{p})\,d\mathbf{p}$ simply denotes the number of molecules per unit volume, having momentum in $d\mathbf{p}$ surrounding $\mathbf{p}$.

(ii) *Boltzmann's Equation*

The (average) number of molecules $f(\mathbf{r}_1,\mathbf{p}_1)\delta\mathbf{r}_1\delta\mathbf{p}_1$ in the element $\delta\mathbf{r}_1\delta\mathbf{p}_1$ surrounding $\mathbf{r}_1,\mathbf{p}_1$ changes as a result of the molecular motion, the external forces $\mathbf{X}$ and the intermolecular collisions:

(i) The change of $f(\mathbf{r}_1,\mathbf{p}_1)\delta\mathbf{r}_1\delta\mathbf{p}_1$ during the finite time interval δt, resulting from molecular motion is $-(\mathbf{c}_1.\partial f/\delta\mathbf{r}_1)\delta\mathbf{r}_1\delta\mathbf{p}_1\delta t$, where $\mathbf{c}_1=\mathbf{p}_1/m$ and $\partial/\delta\mathbf{r}_1$ indicates the gradient.

(ii) The change resulting from the acceleration due to the external forces is $-(\mathbf{X}_1.\partial f/\delta\mathbf{p}_1)\delta\mathbf{r}_1\delta\mathbf{p}_1\delta t$.

(iii) The change of $f(\mathbf{r}_1,\mathbf{p}_1)\delta\mathbf{r}_1\delta\mathbf{p}_1$ resulting from collisions can be divided into a decrease and an increase during δt. The decrease during δt, resulting from collisions of, for instance molecules in $\delta\mathbf{r}_2\delta\mathbf{p}_2$, surrounding $\mathbf{r}_2$, $\mathbf{p}_2$ at the beginning of the time interval, is assumed to be given by $a_{12}f(\mathbf{r}_1,\mathbf{p}_1)f(\mathbf{r}_2,\mathbf{p}_2)\delta\mathbf{r}_1\delta\mathbf{p}_1\delta\mathbf{r}_2\delta\mathbf{p}_2\delta t$, where a_{12} is a quantity with the dimensions of a frequency depending on $|\mathbf{p}_2-\mathbf{p}_1|$ and the geometry of the collision ("*Hypothesis of molecular chaos*" (Jeans 1925, § 65–70); "*Stosszahlansatz*" *of Boltzmann* (Boltzmann 1896/98)). The increase during δt resulting from inverse collisions, for instance between two molecules, initially in $\delta\mathbf{r}_1'\delta\mathbf{p}_1'$ and $\delta\mathbf{r}_2'\delta\mathbf{p}_2'$ and at the end of the collision in $\delta\mathbf{r}_1\delta\mathbf{p}_1$ and $\delta\mathbf{r}_2\delta\mathbf{p}_2$ respectively, is according to the same collision number hypothesis given by $a_{21}f(\mathbf{r}_1',\mathbf{p}_1')f(\mathbf{r}_2',\mathbf{p}_2')\delta\mathbf{r}_1'\delta\mathbf{p}_1'\delta\mathbf{r}_2'\delta\mathbf{p}_2'\delta t$. Application of the laws of mechanics shows that $a_{21}=a_{12}$ (*principle of dynamic reversibility*) and $\delta\mathbf{r}_1'\delta\mathbf{p}_1'\delta\mathbf{r}_2'\delta\mathbf{p}_2'=\delta\mathbf{r}_1\delta\mathbf{p}_1\delta\mathbf{r}_2\delta\mathbf{p}_2$ (*principle of Liouville*). Addition of both these changes and summation over $\delta\mathbf{r}_2$ and $\delta\mathbf{p}_2$ then gives the total change of $f(\mathbf{r}_1,\mathbf{p}_1)\delta\mathbf{r}_1\delta\mathbf{p}_1$ during δt due to collisions.

Therefore the total change of f is given by the following *Boltzmann equation*:

$$\frac{\partial f}{\partial t}=-\mathbf{c}_1.\frac{\partial f}{\delta\mathbf{r}_1}-\mathbf{X}_1.\frac{\partial f}{\delta\mathbf{p}_1}+\iint 2\pi b\,db|\,\mathbf{c}_1-\mathbf{c}_2|\,[f(\mathbf{p}_1')f(\mathbf{p}_2')-f(\mathbf{p}_1)f(\mathbf{p}_2)]\,d\mathbf{p}_2, \qquad \ldots\ldots(2.1)$$

where the summation over $\delta\mathbf{p}_2$ and $\delta\mathbf{r}_2$ is replaced by an integration over $\mathbf{p}_2$ and the collision parameter b of the molecule 2 with respect to 1. The dependence of f on $\mathbf{r}$ is assumed to be so small, that in all terms f can be taken at the same place $\mathbf{r}$. The difference quotient $\delta f/\delta t$ is replaced by $\partial f/\partial t$, which shows that $\partial f/\partial t$ must be considered in a certain sense as a smoothed-out time average over time periods long compared with the intermolecular collision-times. This usual derivation of Boltzmann's equation remains in some points rather obscure, especially in its mixing up "coarse grained" and "fine grained" time averages (cf. Kirkwood 1946).

The collision hypothesis (hypothesis of molecular chaos) on which the derivation is based, limits the applicability of Boltzmann's equation (2.1) to dilute gases in which multiple collisions are a rare occurrence compared with the binary collisions. So the treatment of the transport phenomena, as reported for instance by Chapman and Cowling (1939), can be applied only to gases of normal pressures, and an extension of the theory to liquids or fluids has to start by revising this fundamental Boltzmann's equation. Before considering attempts and advances in this direction it is necessary to mention briefly the problems which have arisen from the application of (2.1) to the so-called H-theorem.

(iii) *H-theorem*

As is well known the equilibrium condition $\partial f/\partial t = 0$ of a fluid under no external field ($\mathbf{X} = 0$, $\partial f/\partial \mathbf{r} = 0$), is satisfied by the equation

$$f(\mathbf{p}_1)f(\mathbf{p}_2) = f(\mathbf{p}_1')f(\mathbf{p}_2') \qquad \ldots\ldots(2.2)$$

(*principle of detailed balance*) being a sufficient, but *not* a necessary, condition for equilibrium. This second conclusion can be drawn by considering Boltzmann's H-function:

$$H = \Sigma_i f(\mathbf{r}_i, \mathbf{p}_i) \ln f(\mathbf{r}_i, \mathbf{p}_i) \delta\mathbf{r}_i \delta\mathbf{p}_i, \qquad \ldots\ldots(2.3\,a)$$

or, assuming that $f(\mathbf{r}_i, \mathbf{p}_i)$ is a nearly continuous function in μ-space:

$$H = \iint f(\mathbf{r}, \mathbf{p}) \ln f(\mathbf{r}, \mathbf{p})\, d\mathbf{r}\, d\mathbf{p}. \qquad \ldots\ldots(2.3\,b)$$

The change of this H-function with time, making use of (2.1) and the fact that the molecular current density at the walls of the vessel is zero, is given by

$$\frac{\partial H}{\partial t} = \frac{1}{4}\int\int\int\int \ln\frac{f(\mathbf{p}_1)f(\mathbf{p}_2)}{f(\mathbf{p}_1')f(\mathbf{p}_2')}\,[f(\mathbf{p}_1')f(\mathbf{p}_2') - f(\mathbf{p}_1)f(\mathbf{p}_2)]\,|\,\mathbf{c}_1 - \mathbf{c}_2\,|\,2\pi b\, db\, d\mathbf{r}_1\, d\mathbf{p}_1\, d\mathbf{p}_2. \qquad \ldots\ldots(2.4)$$

As the integrand of this integral is essentially negative or zero, H can never increase. In equilibrium, when $\partial H/\partial t$ must be zero, $f(\mathbf{p})$ must necessarily satisfy the equation (2.2) expressing detailed balance, from which conclusion the distribution function

$$f(\mathbf{p}) = n(2\pi mkT)^{-3/2} \exp - \frac{\mathbf{p}^2}{2mkT}, \qquad \ldots\ldots(2.5)$$

of Maxwell can easily be derived.

This *H-theorem of Boltzmann* states that when a system starts initially in some non-equilibrium condition, and moves by successive re-distribution of momenta and coordinates in the direction of equilibrium, the function H of Boltzmann gradually decreases until in the final state the equilibrium value of H is reached.

The apparent paradox between the fact that H cannot increase with time, and that the mechanical laws on which the derivation seems to be based are reversible in time, has given rise to much discussion, for a detailed treatment of which the reader is referred to well-known textbooks (P. and T. Ehrenfest 1911, §§ 6, 7, Tolman 1938, Chap. VI). The solution is found in the fact, that the derivation is not only based on the principle of dynamic reversibility, but also on statistical assumptions of a probability character, i.e. the collision number hypothesis. This collision number hypothesis can therefore be said to be true only in an overwhelming number of cases, but nevertheless a small probability of deviation remains. The fact that the number of particles is so very large, however, makes the probability of deviation from the most probable value extremely small.

§ 3. METHOD OF STATISTICAL MECHANICS

(i) *Probability Density of Ensemble in γ-space*

A more detailed discussion of the probability laws governing the change of the distribution function, with time on the basis of the method of § 2 leads to assumptions about the *a priori* probabilities of events which are unusual elements in the mechanical treatment of the phenomena in many particle systems considered here. To avoid this difficulty and to put the theory again on an exact basis, Boltzmann and Gibbs introduced the method of statistical mechanics. This method

considers not one, but a large collection of identical and independent systems, all in different phases, and tries to formulate the *laws which connect the average behaviour of this ensemble of systems with the behaviour of the one system of interest*. In general it can be said that only in the case of equilibrium a satisfying justification of this description of the real system by the average behaviour of an ensemble of systems has been given (§4(i)). For non-equilibrium problems much work has still to be done and only a small number of more or less successful attempts of recent years will be mentioned (§4(ii)).

In the method of statistical mechanics the "state" or "phase" of each system of the ensemble is represented by *one* phase point in the $6N$-dimensional γ-space, with coordinates $\mathbf{r}_1 \ldots \mathbf{r}_N$, $\mathbf{p}_1 \ldots \mathbf{p}_N$. These sets of position and momentum vectors of all molecules together form two vectors in the $3N$-dimensional configurational and momentum subspaces, which will be denoted by $\mathbf{r}^N$ and $\mathbf{p}^N$. Infinitely small volume elements in γ-space: $d\mathbf{r}_1 \ldots d\mathbf{r}_N$, $d\mathbf{p}_1 \ldots d\mathbf{p}_N$ will be denoted by $d\mathbf{r}\, d\mathbf{p}^N$. As each system is represented by one phase point in γ-space the ensemble of a very large number, say Γ, of systems is represented in γ-space by a cloud of Γ phase points. The number density of phase points in γ-space is represented by $\rho(\mathbf{r}^N, \mathbf{p}^N)$, which means that the number of phase points in the element $d\mathbf{r}^N\, d\mathbf{p}^N$ surrounding $\mathbf{r}^N$, $\mathbf{p}^N$ in γ-space is equal to $\rho(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N\, d\mathbf{p}^N$. Evidently $\int\int \rho(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N\, d\mathbf{p}^N = \Gamma$. The quotient of this density in phase space and the total number of systems of the ensemble is called the *probability density* in phase space, which will frequently be denoted by "probability":

$$P_N(\mathbf{r}^N, \mathbf{p}^N) = \rho(\mathbf{r}^N, \mathbf{p}^N)/\Gamma.$$

Thus the probability density satisfies the normalizing condition

$$\int\int P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N\, d\mathbf{p}^N = 1. \qquad \ldots\ldots(3.1)$$

Then the probability of finding one system of the ensemble chosen at random, in a state $d\mathbf{r}^N\, d\mathbf{p}^N$ surrounding $\mathbf{r}^N$, $\mathbf{p}^N$ is given by $P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N\, d\mathbf{p}^N$.

Similarly the probability density $P_h(\mathbf{r}^\eta, \mathbf{p}^\eta)$ is defined to find, in any system of the ensemble chosen at random, a definite set of h specified molecules $\iota, \kappa, \ldots, \lambda$, denoted by $\{\eta\}$, in the state $\mathbf{r}_\iota \ldots \mathbf{r}_\lambda$, $\mathbf{p}_\iota \ldots \mathbf{p}_\lambda$, which can be denoted by $\mathbf{r}^\eta$, $\mathbf{p}^\eta$. Thus $\mathbf{r}^\eta$ and $\mathbf{p}^\eta$ are vectors in the $3h$-dimensional configuration and momentum subspace of the set of molecules $\{\eta\}$. Evidently

$$P_h(\mathbf{r}^\eta, \mathbf{p}^\eta) = \int\int P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^{N-\eta}\, d\mathbf{p}^{N-\eta}, \qquad \ldots\ldots(3.2)$$

where $d\mathbf{r}^{N-\eta}$ and $d\mathbf{p}^{N-\eta}$ represent phase elements of the $3(N-h)$ dimensional subspaces of all molecules but the set of h molecules $\{\eta\}$.

Integration over the momenta gives the probability that one system of the ensemble chosen at random is in a configuration $\mathbf{r}^N$ or $\mathbf{r}^\eta$ as

$$P_N(\mathbf{r}^N) = \int P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{p}^N, \qquad \ldots\ldots(3.3)$$

$$P_h(\mathbf{r}^\eta) = \int P_h(\mathbf{r}^\eta, \mathbf{p}^\eta)\, d\mathbf{p}^\eta = \int P_N(\mathbf{r}^N)\, d\mathbf{r}^{N-\eta}. \qquad \ldots\ldots(3.4)$$

All these functions $P_N(\mathbf{r}^N, \mathbf{p}^N)$, $P_h(\mathbf{r}^\eta, \mathbf{p}^\eta)$, $P_N(\mathbf{r}^N)$, $P_h(\mathbf{r}^\eta)$ will be called *distribution functions*. If the basic assumption of statistical mechanics is correct and the ensemble represents the system under consideration, they describe the average distribution of the molecules of *one* system over the momenta and coordinates. If necessary, these distribution functions will be provided by the adjective "*specific*" to indicate the fact that a *distribution of specified molecules* is considered.

(ii) *Change with Time and Equilibrium Ensembles*

The change of each of the systems of the ensemble with time is governed by the laws of mechanics for molecule ι

$$\dot{\mathbf{p}}_\iota = \frac{\partial H(\mathbf{r}^N, \mathbf{p}^N)}{\partial \mathbf{r}_\iota}, \qquad \dot{\mathbf{r}}_\iota = -\frac{\partial H(\mathbf{r}^N, \mathbf{p}^N)}{\partial \mathbf{p}_\iota}, \qquad \ldots\ldots(3.5)$$

where $H(\mathbf{r}^N, \mathbf{p}^N)$ is the Hamiltonian of the system of molecules. The solution of these equations is in principle given by a set of $6N$ constants of motion, or first integrals:

$$\left.\begin{aligned} I_1(\mathbf{r}^N, \mathbf{p}^N) &\equiv H(\mathbf{r}^N, \mathbf{p}^N) = E = \text{const.}, \\ I_\gamma(\mathbf{r}^N, \mathbf{p}^N) &= \text{const.} \qquad (\gamma = 2, \ldots, 6N-1), \\ I_{6N}(\mathbf{r}^N, \mathbf{p}^N) &= \text{const.} + t. \end{aligned}\right\} \qquad \ldots\ldots(3.6)$$

The first $6N-1$ constants determine the curve of the system in γ-space, whereas I_{6N} determines the time scale along this path-curve. Accordingly for each of the molecules $\partial\dot{\mathbf{p}}_\iota/\partial\mathbf{p}_\iota + \partial\dot{\mathbf{r}}_\iota/\partial\mathbf{r}_\iota = 0$, so that also

$$\Sigma_\iota \left(\frac{\partial \dot{\mathbf{p}}_\iota}{\partial \mathbf{p}_\iota} + \frac{\partial \dot{\mathbf{r}}_\iota}{\partial \mathbf{r}_\iota} \right) = 0. \qquad \ldots\ldots(3.7)$$

This equation shows that the motion of the phase points in γ-space is divergence-free, which means that *a definite volume element in γ-space, which moves with the phase points contained in it through the phase space, keeps the same volume*, although in general its form changes considerably. This *theorem* of Liouville can also be expressed in a different form: as all the phase points (each representing one system) which start inside the volume-element in γ-space, remain inside this element all the time, the *density of phase points remains constant in a volume element which moves with the phase points* in γ-space according to the laws of mechanics:

$$D\rho(\mathbf{r}^N, \mathbf{p}^N)/Dt = 0.$$

Division by the constant number of systems of the ensemble then gives $DP_N/Dt = 0$ or

$$\frac{\partial P_N}{\partial t} = -\Sigma_\iota \mathbf{c}_\iota \cdot \frac{\partial P_N}{\partial \mathbf{r}_\iota} - \Sigma_\iota (\mathbf{X}_\iota + \mathbf{F}_\iota) \cdot \frac{\partial P_N}{\partial \mathbf{p}_\iota}, \qquad \ldots\ldots(3.8)$$

where, as before, $\mathbf{X}_\iota$ denotes the external force and $\mathbf{F}_\iota$ the force due to the other molecules, exerted on the molecule ι. Assuming that the total potential energy of the system is composed of the sum of contributions $\frac{1}{2}\Sigma_{\iota,\kappa}\phi_{\iota\kappa}$, where $\phi_{\iota\kappa} = \phi(r_{\iota\kappa})$ is the potential energy of a pair of two molecules ι and κ, the force on ι is equal to $\mathbf{F}_\iota = -\Sigma_\kappa \partial\phi_{\iota\kappa}/\partial\mathbf{r}_\iota$.

The fundamental problem of statistical mechanics is now how to calculate the actual properties of a real system from the properties of the ensemble of systems moving in γ-space. In the case of equilibrium physicists believe the problem to be solved with sufficient accuracy, although one must confess that a rigorous proof has never been given.

The fundamental hypothesis is *that the average value in the real system of some physical quantity G, which is a function $G(\mathbf{r}^N, \mathbf{p}^N)$ of the momenta and coordinates of all molecules of the system, can be obtained from the representative ensemble by calculating the average value of the function $G(\mathbf{r}^N, \mathbf{p}^N)$, averaged over all systems of the ensemble.* Mathematically this average is given by

$$\overline{G} = \int\int G(\mathbf{r}^N, \mathbf{p}^N) P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N\, d\mathbf{p}^N. \qquad \ldots\ldots(3.9)$$

Evidently an ensemble which describes a system in equilibrium must be stationary, that is, the density $P_N(\mathbf{r}^N, \mathbf{p}^N)$ at a certain point in γ-space, which in general changes with time, has to remain constant. Because of Liouville's theorem, which states that P_N remains constant in a phase element moving with the phase points through γ-space, it is necessary to choose the ensemble initially in such a way that $P_N(\mathbf{r}^N, \mathbf{p}^N)$ is constant along the whole orbit of each system in the γ-space. Thus P_N should be a function $P_N(E, I_2, \ldots, I_{6N-1})$ of the $6N-1$ constants of motion.

The simplest ensemble satisfying this condition, which represents a system in equilibrium, is the "*microcanonical ensemble*" of Boltzmann, filling homogeneously the space in γ-space between two successive energy surfaces $E =$ constant and $E + dE =$ constant with constant probability density P_0.

If this shell degenerates into a surface $E =$ constant by letting dE approach to zero and P_0 to infinity, one introduces a surface probability density $\sigma = A/|\operatorname{grad} E|$ where the constant $A = P_0\, dE$ is normalized according to $\int \sigma\, dS = 1$ equivalent to (3.1), dS being the surface element in the surface $E =$ constant. The average value of $G(\mathbf{p}^N, \mathbf{r}^N)$ is then given by

$$\overline{G} = \int G(\mathbf{p}^N, \mathbf{r}^N)\sigma(\mathbf{r}^N, \mathbf{p}^N)\, dS. \qquad \ldots\ldots (3.9\,b)$$

This surface or microcanonical ensemble of Boltzmann has the disadvantage that in practical calculations of the average value of some quantity, the integration over the phase-space has to be limited to the energy surface $E =$ constant. This difficulty can be avoided by using the method of the generating function, better known under the name of the method of Darwin and Fowler (1922).

On the other hand, this difficulty is avoided by using the "*canonical ensemble*" of Gibbs: $P_N(\mathbf{r}^N, \mathbf{p}^N) = A \exp -\beta H(\mathbf{r}^N, \mathbf{p}^N)$, where $H(\mathbf{r}^N, \mathbf{p}^N)$ is the Hamiltonian, $A =$ constant. This ensemble, which is also a stationary ensemble, is distributed continuously over the γ-space. However, since the volume of γ-space contained between two successive energy shells $E =$ constant and $E + dE =$ constant increases with a very high power of E (about $E^{3N/2-1}$), an overwhelming number of systems of the aggregate has an energy practically equal to the average energy $\overline{E}$, the fluctuation $(E-\overline{E})^2/\overline{E}^2$ being of the order $1/N$. The constant β has to be identified with $1/kT$, where T is the temperature and k is Boltzmann's constant. The average of $G(\mathbf{p}^N, \mathbf{r}^N)$ is given by

$$\overline{G} = A \iint G(\mathbf{r}^N, \mathbf{p}^N) \exp -\beta H(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N\, d\mathbf{p}^N \qquad \ldots\ldots (3.9\,c)$$

For all practical purposes of calculating averages, thecanonical esemble of Gibbs can be identified with the microcanonical aggregate of Boltzmann, but it has the advantage of allowing continuous integrations to be carried out over the entire γ-space.

The way in which one interprets this ensemble of Gibbs is now a matter of taste. On the one hand, it can be accepted as a mathematical trick, which allows the physicist to free himself from the annoying limiting condition that the energy should be kept exactly constant; in this sense the canonical ensemble can be considered as a mathematical method analogous to the method of Darwin and Fowler (1922) or that of Dirichlet in mathematics (cf. Ornstein 1908, P. and T. Ehrenfest 1911, § 22 *b*, Pauli 1927, Uhlenbeck 1927). On the other hand, it can be

argued that in practice systems at constant temperature, surrounded by isothermal conditions at temperature T, are not systems of constant energy at all. In this case Gibbs' ensemble is a better representation of the actual system than is the micro-canonical ensemble. For the equilibrium behaviour this difference in concept is of no importance. For the fluctuation phenomena, however, it is of the utmost importance and no longer a matter of taste: the fluctuations of the energy can only be described using the canonical ensemble.

In the theoretical considerations in the next chapter we will make use throughout of the canonical ensemble of Gibbs to evaluate averages for a gas or liquid in equilibrium at a temperature T.

(iii) *Generic and Specific Phase. The μ-space*

In (i) the state or phase of the gas or liquid under consideration was represented by one representative point in the $6N$-dimensional γ-space. When a system of identical molecules or atoms is considered, it has certain advantages to introduce the concept of 6-dimensional μ-space. Each molecule is then specified by one representative point in μ-space and the situation of the gas is sufficiently specified by the cloud of N representative points in μ-space, corresponding to the N molecules of the system. The word "situation" is used instead of the word state or phase, in order to indicate that for knowledge of the state or phase of the gas, it is necessary to know what molecules correspond to the cloud of points in μ-space, whereas this information is not necessary to describe the situation. Gibbs distinguishes this exact state or phase and this so-called situation by the words specific and generic phase. Evidently $N!$ different phases in γ-space correspond to one and the same situation in μ-space.

When the fundamental hypothesis of statistical mechanics is true the quantity $P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N d\mathbf{p}^N$, defined as the probability of finding one system of the aggregate chosen at random in the volume element $d\mathbf{r}^N d\mathbf{p}^N$ surrounding $\mathbf{r}^N, \mathbf{p}^N$ in γ-space, is equal to the probability of finding the representative phase point for one system considered in γ-space in the volume element $d\mathbf{r}^N d\mathbf{p}^N$ surrounding $\mathbf{r}^N, \mathbf{p}^N$. Thus the probability of finding the representative points of the molecules of the system under consideration in the infinitely small volume elements $d\mathbf{r}_1 d\mathbf{p}_1, \ldots, d\mathbf{r}_N d\mathbf{p}_N$ surrounding the position- and momentum-vectors $\mathbf{r}_1\mathbf{p}_1, \ldots, \mathbf{r}_N\mathbf{p}_N$ in μ-space in an arbitrary order is $N!$ times as large and is given by

$$f_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}_1 d\mathbf{p}_1 \ldots d\mathbf{r}_N d\mathbf{p}_N = N!\, P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N d\mathbf{p}^N, \quad \ldots\ldots (3.10)$$

where $(\mathbf{r}^N, \mathbf{p}^N)$ is now used as a short notation for the set of position- and momentum-vectors $\mathbf{r}_1\mathbf{p}_1, \mathbf{r}_2\mathbf{p}_2, \ldots, \mathbf{r}_N\mathbf{p}_N$. This function f_N will be called the *generic* distribution function.

Similarly the probability for a situation of h molecules in the elements $d\mathbf{r}_1 d\mathbf{p}_1, \ldots, d\mathbf{r}_h d\mathbf{p}_h$ surrounding the momentum- and position-vectors $\mathbf{r}_1\mathbf{p}_1, \mathbf{r}_2\mathbf{p}_2, \ldots, \mathbf{r}_h\mathbf{p}_h \equiv (\mathbf{r}^h, \mathbf{p}^h)$, can be found by multiplying $P_h(\mathbf{r}^h, \mathbf{p}^h)\, d\mathbf{r}^h d\mathbf{p}^h$ (the probability of finding the set of h molecules $1, 2, \ldots, h$ in the volume element $d\mathbf{r}^h d\mathbf{p}^h = d\mathbf{r}_1 d\mathbf{p}_1 \ldots d\mathbf{r}_h d\mathbf{p}_h$ in the $6h$-dimensional subspace of γ-space corresponding to this set $\{h\}$) by a factor $N!/(N-h)!$, whic his the number of possible ways in which a sequence of h molecules can be chosen out of N. Thus

$$f_h(\mathbf{r}^h, \mathbf{p}^h) = \frac{N!}{(N-h)!}\, P_h(\mathbf{r}^h, \mathbf{p}^h). \quad \ldots\ldots (3.11)$$

In particular for $h=1$, one obtains $f_1(\mathbf{r}_1, \mathbf{p}_1)=NP_1(\mathbf{r}_1, \mathbf{p}_1)$, which is immediately clear. In consequence of the relation (3.2) which exists between P_h and P_N, the following relation exists between f_h and f_N:

$$(N-h)!f_h(\mathbf{r}^h, \mathbf{p}^h)=\int\ldots.\int f_N(\mathbf{r}^N, \mathbf{p}^N)\,d\mathbf{r}_{h+1}d\mathbf{p}_{h+1}\ldots.\,d\mathbf{r}_N\,d\mathbf{p}_N. \quad \ldots\ldots (3.12)$$

Similar equations hold for the " generic " configurational distribution functions, which can now easily be summarized:

$$n_h(\mathbf{r}^h)=\frac{N!}{(N-h)!}P_h(\mathbf{r}^h), \quad \ldots\ldots (3.13)$$

$$(N-h)!\;n_h(\mathbf{r}^h)=\int\ldots.\int n_N(\mathbf{r}^N)\,d\mathbf{r}_{h+1}\ldots.\,d\mathbf{r}_N. \quad \ldots\ldots (3.14)$$

The use of these generic distribution functions $f_h(\mathbf{r}^h, \mathbf{p}^h)$ and $n_h(\mathbf{r}^h)$ instead of $P_h(\mathbf{r}^n, \mathbf{p}^n)$ and $P_h(\mathbf{r}^n)$ has the advantage of the disappearance of frequently occurring factors $N!$ and $(N-h)!$ in the different expressions. For that reason these generic distribution functions have been used by Yvon (1935) and by Born and Green (1946, 1947).

(iv) *The Probability Density Matrix in Quantum Statistical Mechanics*

In quantum mechanics it is impossible to specify the state of a system by giving the exact values of all the coordinates and momenta. Instead of this classical method the state of the system is specified by giving the wave function $\psi(\mathbf{r}^N, t)$, a function of all the coordinates and of time, the future development of which is determined by the Schrödinger wave equation. If the wave function is normalized to unity, as we shall always assume, the square $|\psi(\mathbf{r}^N, t)|^2$ gives the probability density of finding the system in the configuration $\mathbf{r}^N$ at a time t. It need not be added, that the word probability used here has nothing to do with the statistical methods introduced in statistical mechanics, its origin being of quantum mechanical nature.

It is often of advantage to make an eigenfunction expansion of the state $\psi(\mathbf{r}^N, t)$ in terms of a complete orthogonal and normalized system of eigenfunctions $\psi_\varrho(\mathbf{r}^N)$, as for instance the eigenfunctions of the energy $\psi_\sigma(\mathbf{r}^N)$ with the set of quantum numbers $\{\sigma\}=\sigma_1, \ldots, \sigma_{3N}$. Then

$$\psi(\mathbf{r}^N, t)=\Sigma_\varrho c_\varrho(t)\psi_\varrho(\mathbf{r}^N). \quad \ldots\ldots (3.15)$$

Because of the normalization of ψ, $\int|\psi(\mathbf{r}^N, t)|^2\,d\mathbf{r}^N=\Sigma_\varrho|c_\varrho|^2=1$, and so $|c_\varrho(t)|^2$ gives the probability of finding the system in the state $\{\rho\}$, and the summation is carried out over all different states $\{\rho\}$.

The knowledge of these probability coefficients $c_\varrho(t)$ is entirely equivalent to the knowledge of the wave function $\psi(\mathbf{r}^N, t)$. As is well known the future behaviour of these probability coefficients is determined completely by the "generalized" or "transformed" Schrödinger equation. Thus this description in terms of the probability coefficients offers a general representation freed from the special rôle played by the coordinates.

Of course the values of the probability coefficients depend entirely on the particular system of orthonormal functions which is used as a base for the description. A change of this basic system to another system also changes the probability coefficients into new ones. If this change is defined by the unitary matrix $\alpha_{\varrho\tau}$,

for which $\alpha^*_{r\varrho} = \alpha^{-1}_{\varrho r}$, the change of the orthonormal functions is given by $\chi_r(\mathbf{r}^N) = \Sigma_\varrho \phi_\varrho(\mathbf{r}^N)\alpha_{\varrho r}$ and the coefficients d_r of $\psi = \Sigma_r d_r \chi_r$ by $d_r = \Sigma_\varrho \alpha^{-1}_{r\varrho} c_\varrho$ respectively.

Thus far we have discussed the quantum mechanical treatment of *one* system. The calculation of the future development of a system is rather cumbersome when the system consists of a large number of particles. The difficulties here are of the same nature as those which occurred in the classical treatment. Therefore here again the concept is introduced of an ensemble of identical systems distributed in a definite way over the different states. The probability of finding a system chosen at random in different states is now specified by a probability matrix representing the average value over the ensemble of either $\psi^*(\mathbf{r}^{N\prime}, t)\psi(\mathbf{r}^N, t)$ or $c^*_{\varrho'}(t)c_\varrho(t)$, both averages being connected by the equation obtained from (3.15):

$$\overline{\psi^*(\mathbf{r}^{N\prime}, t)\psi(\mathbf{r}^N, t)} = \Sigma_{\varrho'}\Sigma_\varrho \overline{c^*_{\varrho'}(t)c_\varrho(t)}\phi^*_{\varrho'}(\mathbf{r}^{N\prime})\phi_\varrho(\mathbf{r}^N). \quad \ldots\ldots(3.16)$$

It has been shown by von Neumann (1927, 1932b) and Dirac (1929, 1930, 1935) that the rôle of the probability density $P_N(\mathbf{r}^N, \mathbf{p}^N)$ of classical statistical mechanics is taken over in quantum statistical mechanics by the Hermitian *probability density matrix* which can be defined either by the average value

$$\mathscr{P}_N(\mathbf{r}^N;\ \mathbf{r}^{N\prime}) = \overline{\psi^*(\mathbf{r}^{N\prime}, t)\psi(\mathbf{r}^N, t)}, \quad \ldots\ldots(3.17)$$

or in the generalized language by the average value

$$\mathscr{P}_{\varrho\varrho'} = \overline{c^*_{\varrho'}(t)c_\varrho(t)}. \quad \ldots\ldots(3.18)$$

In this second form $\mathscr{P}_{\varrho\varrho'}$ depends on the particular system of wave functions $\phi_\varrho(\mathbf{r}^N)$ which is chosen as a basis for the description of the system. A transition to a new system $\chi_r = \Sigma_\varrho \phi_\varrho \alpha_{\varrho r}$ gives for the new system

$$\mathscr{P}_{rr'} = \Sigma_\varrho \Sigma_{\varrho'} \alpha^{-1}_{r\varrho} \mathscr{P}_{\varrho\varrho'} \alpha_{\varrho' r'} = \Sigma_\varrho \Sigma_{\varrho'} \mathscr{P}_{\varrho\varrho'} \alpha^*_{\varrho r} \alpha_{\varrho' r'}$$

according to the well-known laws of transformation theory. If δ functions ("coordinate eigenfunctions") are chosen as a basis for the description of the system, $\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^{N\prime})$ can also be considered as a special form of the generalized density matrix $\mathscr{P}_{\varrho\varrho'}$. According to (3.16) the relation between the two forms is

$$\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^{N\prime}) = \Sigma_\varrho \Sigma_{\varrho'} \mathscr{P}_{\varrho\varrho'} \phi^*_{\varrho'}(\mathbf{r}^{N\prime})\phi_\varrho(\mathbf{r}^N). \quad \ldots\ldots(3.19)$$

From the definition (3.17) it can be seen that $\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^{N\prime})$ does *not* depend on any particular system of wave functions, since it is the average value $\overline{\psi^*(\mathbf{r}^{N\prime})\psi(\mathbf{r}^N)}$ over all systems of the ensemble. *The expression* (3.19) *for* $\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^{N\prime})$ *is indeed independent of a change of the basic system of eigenfunctions* from $\phi_\varrho(\mathbf{r}^N)$ to, for instance, $\chi_r(\mathbf{r}^N)$, which follows immediately from the transformation formula for $\mathscr{P}_{\varrho\varrho'}$ and the relations $\chi_r = \Sigma\phi_\varrho\alpha_{\varrho r}$ connecting ϕ_ϱ and χ_r. This invariance of the probability density matrix is very useful for actual calculations.

To these matrix elements can be correlated a "*probability operator*" $\mathscr{P}$, which allows the matrix elements to be written as

$$\mathscr{P}_N(\mathbf{r}^N;\ \mathbf{r}^{N\prime}) = \Sigma_\varrho \phi^*_\varrho(\mathbf{r}^{N\prime})\mathscr{P}\phi_\varrho(\mathbf{r}^N), \quad \ldots\ldots(3.20)$$

$$\mathscr{P}_{\varrho\varrho'} = \int \phi^*_\varrho(\mathbf{r}^N)\mathscr{P}\phi_{\varrho'}(\mathbf{r}^N)\,d\mathbf{r}^N. \quad \ldots\ldots(3.21)$$

The choice of this probability operator $\mathscr{P}$ and therefore of the probability density matrix $\mathscr{P}_{\varrho\varrho'}$ or $\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^{N\prime})$ depends on the system which has to be represented by the ensemble. It will be shown later in this section that for an equilibrium ensemble the operator $\mathscr{P}$ has to be a function of the Hamilton operator alone.

The rôle of the probability density $P_N(\mathbf{r}^N, \mathbf{p}^N)$ of classical statistical mechanics, being a function of the set of N position vectors $\mathbf{r}^N$ and of N momentum vectors $\mathbf{p}^N$, is taken over in quantum theory by this probability density matrix $\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^{N\prime})$, which is a function of the two sets of N position vectors $\mathbf{r}^N$ and $\mathbf{r}^{N\prime}$ (or in general by $\mathscr{P}_{\varrho\varrho'}$, which is a function of both sets of quantum numbers $\{\xi\}$ and $\{\xi'\}$). The integration over the momenta in classical theory, which leads to the probability density $P_N(\mathbf{r}^N) = \int P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{p}^N$ of finding a system of the ensemble chosen at random in the state $\mathbf{r}^N$, is replaced in quantum theory by the taking of the diagonal term $\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^N)$. This is in fact the quantum mechanical expression for the probability density to find a system chosen at random in the state $\mathbf{r}^N$.

The probability density matrix is always taken to be normalized, in similar manner to the probability density in classical theory (equation (3.1)), according to the relation

$$\int \mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^N)\, d\mathbf{r}^N = 1. \qquad \ldots\ldots (3.22)$$

In analogy to the reduced probability densities $P_h(\mathbf{r}^\eta, \mathbf{p}^\eta)$ introduced in section (i) giving the probability density for the occurrence of position vectors $\mathbf{r}^\eta$ and momentum vectors $\mathbf{p}^\eta$ of the h molecules of the set $\{\eta\} \equiv \iota, \varkappa \ldots . \nu$, it is possible to introduce reduced probability density matrices $\mathscr{P}_h(\mathbf{r}^\eta;\ \mathbf{r}^{\eta\prime})$ of this same set of h molecules. This reduced matrix is obtained by taking the diagonal elements corresponding to the molecules *not* occurring in the group $\{\eta\}$ and integrating over the corresponding position vectors

$$\mathscr{P}_h(\mathbf{r}^\eta;\ \mathbf{r}^{\eta\prime}) = \int \mathscr{P}(\mathbf{r}^\eta \mathbf{r}^{N-\eta};\ \mathbf{r}^{\eta\prime}\mathbf{r}^{N-\eta})\, d\mathbf{r}^{N-\eta}. \qquad \ldots\ldots (3.23)$$

These reduced density matrices introduced originally by Husimi (1940), who studied their properties, have been used especially by Born and Green (1947 c) in their recent attempt to give the foundation of quantum hydrodynamics on quantum mechanics. The diagonal term $\mathscr{P}_h(\mathbf{r}^\eta;\ \mathbf{r}^\eta)$ is the quantum mechanical analogue of the classical reduced distribution function $P_h(\mathbf{r}^\eta)$.

Here also the normalization of the density matrix can be chosen in analogy with the "generic distribution functions" introduced in section (iii). The probability density matrix, which is the quantum statistical analogue of the generic probability density $f_h(\mathbf{r}^h, \mathbf{p}^h)$ is, according to (3.11),

$$\mathscr{N}(\mathbf{r}^h;\ \mathbf{r}^{h\prime}) = \frac{N!}{(N-h!)}\, \mathscr{P}(\mathbf{r}^h;\ \mathbf{r}^{h\prime}), \qquad \ldots\ldots (3.24)$$

and the diagonal elements of this matrix are the analogues of the generic configurational probability densities

$$\mathscr{N}(\mathbf{r}^h;\ \mathbf{r}^h) = \frac{N!}{(N-h)!}\, \mathscr{P}(\mathbf{r}^h;\ \mathbf{r}^h), \qquad \ldots\ldots (3.25)$$

which were introduced by (3.13).

(v) *Time Dependence and Equilibrium Ensembles in Quantum Statistical Mechanics*

The time dependence of the probability density matrix can immediately be derived from the time dependence of ψ according to the Schrödinger equation $\partial\psi/\partial t = -(i/\hbar)\mathscr{H}\psi$, where $\mathscr{H}$ is the Hamilton operator, or from the time dependence of c_ϱ according to the Schrödinger equation $\partial c_\varrho/\partial t = -(i/\hbar)\Sigma_{\varrho'}\mathscr{H}_{\varrho\varrho'}c_{\varrho'}$. This gives directly

$$\frac{\partial \mathscr{P}_{\varrho\varrho'}}{\partial t} = -\frac{i}{\hbar}(\mathscr{H}\mathscr{P} - \mathscr{P}\mathscr{H})_{\varrho\varrho'}, \qquad \ldots\ldots(3.26)$$

or

$$\frac{\partial \mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^{N\prime})}{\partial t} = -\frac{i}{\hbar}(\mathscr{H}\mathscr{P} - \mathscr{P}\mathscr{H})(\mathbf{r}^N;\ \mathbf{r}^{N\prime}). \qquad \ldots\ldots(3.27)$$

These equations have to be taken in the place of the classical Liouville equation (3.8), which can be seen by writing the right-hand side of the classical Liouville equation (3.8) in the form of the Poisson bracket

$$-\{H, P_N\} = -\Sigma_\iota[(\partial H/\partial \mathbf{p}_\iota).(\partial P_N/\partial \mathbf{r}_\iota) - (\partial H/\partial \mathbf{r}_\iota).(\partial P_N/\partial \mathbf{p}_\iota)],$$

of which the commutator multiplied by $(i/\hbar)$ of equation (3.27) is the quantum mechanical analogue.

For equilibrium it is now required that $\mathscr{P}$ is a function of the Hamilton operator $\mathscr{H}$. As in classical statistical mechanics, we shall make use only of the operator in the form $\mathscr{P} \propto \exp -\beta\mathscr{H}$, where $\exp -\beta\mathscr{H}$ is defined to be $\Sigma_k(-\beta\mathscr{H})^k/k!$. The ensemble for which the operator has this particular form will be called the canonical ensemble.

This probability density matrix can now be used to calculate averages in the aggregate of systems. If the system is in a state specified by the wave function $\psi(\mathbf{r}^N)$, the result of an experimental determination of the value of a physical quantity $G(\mathbf{r}^N, \mathbf{p}^N)$ which is a function of coordinates and momenta can only be expressed in the form of an expectation value G. This expectation value is given by $G_\psi = \int \psi^* \mathscr{G}\psi\, d\mathbf{r}^N$, where $\mathscr{G}$ is a Hermitian operator obtained by replacing in $G(\mathbf{r}^N, \mathbf{p}^N)$ everywhere the momentum $\mathbf{p}_\iota$ of the molecule ι by $p_\iota = (\hbar/i)\partial/\partial \mathbf{r}_\iota$. This procedure which is essential for the treatment of the system by quantum mechanics, has of course nothing to do with the statistical methods which are needed because of the large number of molecules. The averaging process of G over all systems of the aggregate gives

$$\overline{G} = \overline{\int \psi^*(\mathbf{r}^N)\mathscr{G}\psi(\mathbf{r}^N)\, d\mathbf{r}^N} = \Sigma_\varrho \Sigma_{\varrho'} \overline{c_\varrho^* c_{\varrho'}} \mathscr{G}_{\varrho'\varrho} \qquad \ldots\ldots(3.28)$$

where $\mathscr{G}_{\varrho\varrho'}$ is the matrix element $\int \phi_\varrho^* \mathscr{G}\phi_{\varrho'}\, d\mathbf{r}^N$ of G. Using the definition of the density matrix, this average value of G can be expressed in the form of a trace of either of the two forms of the probability density matrix

$$\overline{G} = \int (\mathscr{P}\mathscr{G})(\mathbf{r}^N;\ \mathbf{r}^N)\, d\mathbf{r}^N = \Sigma_\varrho (\mathscr{P}\mathscr{G})_{\varrho\varrho}, \qquad \ldots\ldots(3.29)$$

where $(\mathscr{P}\mathscr{G})(\mathbf{r}^N;\ \mathbf{r}^{N\prime})$ and $(\mathscr{P}\mathscr{G})_{\varrho\varrho'}$ are matrix elements of the operator $\mathscr{P}\mathscr{G}$ defined according to (3.20) and (3.21). Since these two expressions are the trace of a matrix the operators $\mathscr{P}$ and $\mathscr{G}$ can be reversed in order.

The calculation of the average value of a quantity G is reduced in quantum theory to the integration over $\mathbf{r}^N$ of the quantity $(\mathscr{P}\mathscr{G})(\mathbf{r}^N, \mathbf{r}^N)$, which evidently

takes the place of the classical quantity $\int P_N(\mathbf{r}^N, \mathbf{p}^N) G(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{p}^N$. Thus the following principle can be used to generalize classical theories to make them valid in quantum mechanics: *In all expressions the integrals either of $P_N(\mathbf{r}^N, \mathbf{p}^N)$ or of a product $P_N(\mathbf{r}^N, \mathbf{p}^N).G(\mathbf{r}^N, \mathbf{p}^N)$ over the momentum vectors $\mathbf{p}^N$ can simply be replaced by the quantum mechanical analogue: the diagonal elements of the matrix $\mathscr{P}(\mathbf{r}^N; \mathbf{r}^{N'})$ or of the matrix $(\mathscr{P}\mathscr{G})(\mathbf{r}^N; \mathbf{r}^{N'})$ corresponding to the product PG* (cf. Born and Green 1947 c).

§4. THE PROBLEM OF THE JUSTIFICATION OF THE METHOD OF STATISTICAL MECHANICS

(i) *Equilibrium*

The proof of the fundamental hypothesis of statistical mechanics, stating that the average value in time of a function $G(\mathbf{r}^N, \mathbf{p}^N)$ of the momenta and coordinates of a system in equilibrium

$$\widetilde{G} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau G(\mathbf{r}^N, \mathbf{p}^N)\, dt \qquad \ldots\ldots (4.1)$$

can be calculated by taking the average of $G(\mathbf{r}^N, \mathbf{p}^N)$ over all systems of the ensemble, distributed over the surface of constant energy in phase space

$$\overline{G} = \int \sigma(\mathbf{r}^N, \mathbf{p}^N) G(\mathbf{r}^N, \mathbf{p}^N)\, dS, \qquad \ldots\ldots (4.2)$$

was based by Maxwell and Boltzmann on the celebrated *ergodic hypothesis* or the "*principle of the continuity of path*". As, according to this hypothesis, the phase point, which represents the system, passes sooner or later through every point on the surface of constant energy in γ-space, the surface ensemble of all phase points move in one and the same path curve on the energy surface. The motions of these phase points are only distinguished by the value of the additive constant in the last of the integrals of motion (3.6), so the time average is exactly the same for all systems of the ensemble, which gives $\widetilde{G} = \overline{\widetilde{G}}$. Interchange of the two averaging procedures then gives $\widetilde{\overline{G}} = \overline{\widetilde{G}}$, whereas finally, because of the fact that the ensemble is chosen to be stationary, the average value $\overline{G}$ is independent of time, making $\widetilde{\overline{G}} = \overline{G}$. Thus taken together one finds $\widetilde{G} = \overline{\widetilde{G}} = \widetilde{\overline{G}} = \overline{G}$, which proves the fundamental hypothesis of statistical mechanics (Maxwell 1879, Boltzmann 1871). The justification of the method of statistical mechanics for the surface ensemble of constant energy also justifies the use of the canonical ensemble of Gibbs, which is mathematically equivalent with Boltzmann's surface ensemble of exactly constant energy.

The improbability of the ergodic hypothesis, due to the presence of special, though exceptional, periodic motions, was stressed by Kelvin (1891) and Poincaré (1894), and the proof of its incorrectness finally given by Plancherel and Rosenthal (1913). P. and T. Ehrenfest (1911) replaced this hypothesis by the *quasi-ergodic hypothesis* which stated that the path curve, though not going through every point of the surface, covers the energy surface so called "everywhere dense", whereas the exceptional path curves, which do *not* cover the energy surface, "everywhere dense", are of Lebesgue measure zero (Caratheodory 1919, Birkhoff 1922). They

indicated clearly, however, that no proof of the fundamental hypothesis of statistical mechanics can be based on this quasi-ergodic theorem. A clearer formulation of the situation became possible after the introduction of the notion of *metrical transitivity* by Birkhoff and Smith (1928) and Koopman (1931) a continuous single-parameter group of transformations of phase space in itself having the property of metrical transitivity if, and only if, the phase space cannot be divided into two or more subspaces, each of measure greater than zero, and each invariant under every transformation of the group. v. Neumann (1932a) and Birkhoff (1931, 1932) finally gave the proof *that the fundamental hypothesis of statistical mechanics is true, provided that the transformation in phase space have this property of metrical transitivity*. So the problem of the truth or falsity of the fundamental hypothesis of statistical mechanics is reduced to that of the metrical transitivity of Hamiltonian systems. Although this metrical transitivity of the Hamiltonians of systems which are studied in physics has never been proved in general, most physicists believe that it is actually true.

Therefore in general the average value of a physical quantity for a system containing a large number of molecules can be identified with the average value of this quantity over the ensemble which represents the system. In particular this also holds for the molecular distribution: where f_1 is defined as

$$f_1(\mathbf{r}_1, \mathbf{p}_1) = N \iint P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^{N-1}\, d\mathbf{p}^{N-1}, \qquad \ldots\ldots (4.2\,a)$$

the quantity $f_1(\mathbf{r}_1, \mathbf{p}_1)\, d\mathbf{r}_1\, d\mathbf{p}_1$ being the number of molecules in $d\mathbf{r}_1$, $d\mathbf{p}_1$ averaged over the ensemble according to statistical mechanics, can be identified with $f(\mathbf{r}_1, \mathbf{p}_1)\, d\mathbf{r}_1\, d\mathbf{p}_1$ the average number of molecules in the real system, averaged in time according to the kinetic theory of § 2. In this case the fact that $f(\mathbf{r}_1, \mathbf{p}_1)$ was defined as a function, "coarse grained" in space, with finite volume elements, is not of much importance, because *in the case of equilibrium* the finite elements $d\mathbf{r}_1\, d\mathbf{p}_1$ can be made smaller and smaller, so that the quotient of the total number of molecules $f(\mathbf{p}_1, \mathbf{r}_1)\, d\mathbf{r}_1\, d\mathbf{p}_1$, in $d\mathbf{r}$, $d\mathbf{p}$, and the volume $d\mathbf{r}_1\, d\mathbf{p}_1$ approaches a finite limit, the probability density $f(\mathbf{r}_1, \mathbf{p}_1)$. The quantity $n_1(\mathbf{r}_1) = N \int P_N(\mathbf{r}^N)\, d\mathbf{r}^{N-1}$ becomes identical with the number density $n(\mathbf{r}_1)$ of the real system.

(ii) *Non-equilibrium*

The use of the ensemble method for systems which are *not* in equilibrium has, up till now, only been considered from a far less exact point of view and in fact it is not at all certain with what kind of averaging procedure the physical quantities of a system which is not in equilibrium can be calculated from the probability density of an ensemble. It is not our purpose to treat the existing literature on this point here in detail, this report being concerned mainly with the equilibrium problems of gases, but some points may be of importance.

One of the first problems of non-equilibrium systems, which has been discussed already in much detail by several authors, is the definition in statistical mechanics of the analogue of the H-function of Boltzmann, defined in equation (2.36) in terms of the kinetic distribution function f. In the case of equilibrium this offers no difficulties, the definition of H in statistical mechanics being, according to Gibbs,

$$H = \iint P_N(\mathbf{r}^N, \mathbf{p}^N) \ln P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N\, d\mathbf{p}^N. \qquad \ldots\ldots (4.3)$$

The H-function of Gibbs can thus be interpreted as the ensemble average of the quantity $\ln P_N(\mathbf{r}^N, \mathbf{p}^N)$. For a system of "independent" molecules, where the interaction between the molecules can be neglected,

$$P_N = P_1(\mathbf{r}_1, \mathbf{p}_1) \,.\, \ldots \,.\, P_1(\mathbf{r}_N, \mathbf{p}_N).$$

Substituting this in (4.3) and introducing the generic distribution function $f_1(\mathbf{r}_1, \mathbf{p}_1) = NP_1(\mathbf{r}_1, \mathbf{p}_1)$ one obtains

$$H = \int\int f_1(\mathbf{r}_1, \mathbf{p}_1) \ln f_1(\mathbf{r}_1, \mathbf{p}_1)\, d\mathbf{r}_1\, d\mathbf{p}_1 - N \ln N; \qquad \ldots\ldots (4.4)$$

$f_1(\mathbf{r}_1, \mathbf{p}_1)\, d\mathbf{r}_1\, d\mathbf{p}_1$ is in the equilibrium theory the average number of molecules in $d\mathbf{r}_1\, d\mathbf{p}_1$ and should thus be equal to the kinetic distribution function f defined in the kinetic theory in μ-space. As in the case of equilibrium this identification offers no difficulties, *the H-function of statistical mechanics, being the ensemble average of* $\ln P_N(\mathbf{r}^N, \mathbf{p}^N)$, *becomes* (except for the unimportant additive constant term $N \ln N$) *identical with the kinetically defined H-function* $\int f(\mathbf{r}, \mathbf{p}) \ln f(\mathbf{r}, \mathbf{p})\, d\mathbf{r}\, d\mathbf{p}$. The H-function defined according to (4.3) can be shown to be the correct expression for minus the entropy of the system divided by the Boltzmann constant k.

In the case where the system is *not* in equilibrium, the situation is quite different. Originally Gibbs (1902, Chap. XI) compared the value of H according to (4.3) for the equilibrium distribution, with the value of H for an arbitrary, non-equilibrium distribution but having the *same value* of the average energy. Then Gibbs proved that this non-equilibrium value of H is *larger* than the equilibrium value, so that the H-function of the equilibrium distribution is a minimum compared with non-equilibrium distributions having the same average total energy. Although this suggests that $-Hk$ might give a general statistical definition of the entropy, which also holds for non-equilibrium states (the entropy being a maximum compared with neighbour states of the same energy U), this conclusion is wrong. The difficulty is that, because of the theorem of Liouville, stating that the probability density $P_N(\mathbf{r}^N, \mathbf{p}^N)$ remains constant moving with a representative point through the γ-space, each function of $P_N(\mathbf{r}^N, \mathbf{p}^N)$ integrated over the entire phase space also remains constant. Thus the function $H = \int P_N \ln P_N\, d\mathbf{r}^N\, d\mathbf{p}^N$, which for the non-equilibrium aggregate had a value larger than the equilibrium value of H of the same energy U, *remains constant throughout time and never decreases to approach the equilibrium value*. So the identification of (4.3) with minus entropy is in general impossible.

What really happens during the approach to equilibrium of an aggregate of systems which is not in equilibrium, will be explained using a simple example of a number of phase points which are distributed non-uniformly over the space between two neighbouring energy surfaces in γ-space. The equilibrium distribution, which consists of a *uniform* distribution between these two surfaces and corresponds to a lower value of H, arises from the non-uniform distribution as a result of the "stirring" caused by the mechanical motion of the representative points of the aggregate between these two energy surfaces in γ-space. As in the case of two liquids, blue and yellow, which by stirring are mixed until in the end a liquid results which is green in colour, so finally a uniform distribution of phase points between the energy surfaces results because of the mechanical motion of molecules of the system. In fact, however, existing discontinuities of the distribution of phase points do not disappear: the density round a certain phase point

keeps the same value during the motion, only the form of the element of phase space changes, and so a region of phase space, where originally the probability density was too high, deforms finally into a long and thin thread of the same density, which covers the entire surface. So what can really be said is, that *the average density of phase points between the energy surfaces, averaged over a small but finite space element of γ-space, approaches in the course of time the limiting value, which corresponds to the uniform equilibrium distribution of phase points corresponding to the microcanonical aggregate.* This approach to equilibrium has been considered in more detail by Gibbs (1902, Chap. XII) and, among others, by Lorentz (1907, § 79).

As a result of these considerations Ehrenfest (1906, 1911) introduced the concept of the "*coarse grained*" *probability* in phase space,

$$\bar{P}_N(\mathbf{r}^N, \mathbf{p}^N) = \frac{\overset{(\delta)}{\iint} P_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N\, d\mathbf{p}^N}{\delta\mathbf{r}^N\, \delta\mathbf{p}^N}, \qquad \ldots\ldots(4.5)$$

being the average of P_N over a small but finite region $\delta\mathbf{r}^N\delta\mathbf{p}^N$ in phase space. In terms of these coarse grained probabilities it is possible to give the definition of the function

$$\begin{aligned} H &= \Sigma \bar{P}_N(\mathbf{r}^N, \mathbf{p}^N) \ln \bar{P}_N(\mathbf{r}^N, \mathbf{p}^N) \delta\mathbf{r}^N \delta\mathbf{p}^N \\ &= \iint P_N(\mathbf{r}^N, \mathbf{p}^N) \ln \bar{P}_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N\, d\mathbf{p}^N. \qquad \ldots\ldots(4.6) \end{aligned}$$

This definition of H, being *the ensemble average of the logarithm of the coarse grained probability density, gives the real statistical mechanical expression for minus the entropy divided by k.* Evidently the value of this H-function so defined changes with time: it can be proved that H can only decrease until a uniform distribution corresponding to equilibrium is reached and then becomes constant. Some difficulties still remain (Lorentz 1907, § 79): it is only proved that a non-uniform ensemble of constant energy approaches the microcanonical ensemble, but it cannot be seen that the ensemble approaches a canonical ensemble in the course of time. In practice, however, this will not give rise to difficulties, when one considers the canonical ensemble as a fair approximation of the microcanonical ensemble.

According to these considerations, for a gas of independent molecules, in the general case when the gas is not in equilibrium, the statistical mechanical analogue for the kinetic distribution function $f(\mathbf{r}, \mathbf{p})$ occurring in the H-function $\int f \ln f\, d\mathbf{r}\, d\mathbf{p}$ should be $\bar{f}_1 = N \iint \bar{P}_N\, d\mathbf{r}^{N-1}\, d\mathbf{p}^{N-1}$. This runs parallel with the fact that it seems to be impossible in this non-equilibrium case to define $f(\mathbf{r}, \mathbf{p})$ kinetically without using finite elements of the μ-space!

Recently Kirkwood (1946) made an interesting advance in the problem of finding the type of *averaging procedure for non-equilibrium systems*, which must be used in statistical mechanics of non-stationary ensembles to obtain the experimental average values, by introducing *time-smoothed probability densities* instead of the coarse grained probability densities which were introduced in the preceding considerations.

The time average of a function $G(\mathbf{r}^N, \mathbf{p}^N)$, not explicitly dependent on time, is expressible in the form

$$\tilde{G} = \frac{1}{\tau} \int_0^\tau G(\Omega_s \mathbf{r}^N, \Omega_s \mathbf{p}^N)\, ds, \qquad \ldots\ldots(4.7)$$

where Ω_s is an operator which transforms the coordinates and momenta of the representative point $\mathbf{r}^N, \mathbf{p}^N$ at time t into the coordinates and momenta of this same representative point at a time $t+s$. The ensemble average of this same function $G(\mathbf{r}^N, \mathbf{p}^N)$ can be obtained from the probability density at time t, $P_N(\mathbf{r}^N, \mathbf{p}^N, t)$, by the expression

$$\overline{G} = \int\int G(\mathbf{r}^N, \mathbf{p}^N) P_N(\mathbf{r}^N, \mathbf{p}^N, t)\, d\mathbf{r}^N\, d\mathbf{p}^N, \qquad \ldots\ldots(4.8)$$

where t is kept constant during the integration.

The fundamental assumption which is now introduced by Kirkwood is, that *in time dependent statistical mechanics the experimental value must be identified with the ensemble average of the time average of* $G(\mathbf{r}^N, \mathbf{p}^N)$, viz.:

$$\overline{\widetilde{G}} = \frac{1}{\tau}\int\int\int_0^\tau G(\Omega_s \mathbf{r}^N, \Omega_s \mathbf{p}^N) P_N(\mathbf{r}^N, \mathbf{p}^N, t)\, ds\, d\mathbf{r}^N\, d\mathbf{p}^N. \qquad \ldots\ldots(4.9)$$

This expression can now be transformed using Liouville's theorem:

$$P_N(\mathbf{r}^N, \mathbf{p}^N, t) = P_N(\Omega_s \mathbf{r}^N, \Omega_s \mathbf{p}^N, t+s), \quad d\mathbf{r}^N d\mathbf{p}^N = d\Omega_s \mathbf{r}^N\, d\Omega_s \mathbf{p}^N$$

giving

$$\overline{\widetilde{G}}(t) = \int\int G(\mathbf{r}^N, \mathbf{p}^N) \widetilde{P}_N(\mathbf{r}^N, \mathbf{p}^N, t)\, d\mathbf{r}^N\, d\mathbf{p}^N \qquad \ldots\ldots(4.10)$$

where $\widetilde{P}_N$, the time-smoothed probability density is *defined* by

$$\widetilde{P}_N(\mathbf{r}^N, \mathbf{p}^N, t) = \frac{1}{\tau}\int_0^\tau P_N(\mathbf{r}^N, \mathbf{p}^N, t+s)\, ds. \qquad \ldots\ldots(4.11)$$

So in a system not in equilibrium, according to Kirkwood, the average value of G must be calculated using the time-smoothed average of the probability density of the ensemble taken over a time τ *at fixed values* $\mathbf{p}^N, \mathbf{r}^N$.

It is important to note that, when averages are considered in which the integration is carried out over *all* momenta and coordinates, this expression (4.10) is identical with the time average over the ensemble average of $G(\mathbf{r}^N, \mathbf{p}^N)$. Kirkwood showed that the lack of commutability of both averaging procedures, when the integration is not carried out over all coordinates and momenta, is closely connected with the dissipative processes which occur in the drift towards equilibrium.

Kirkwood first applied the theory to the Brownian motion, being able to derive from the exact equations of statistical mechanics, the equation of Langevin for the motion of a particle in a fluid, in which occur the frictional retarding force proportional to the velocity and the fluctuating force due to the collisions with other molecules. The assumption which was introduced to derive this result was that the time τ over which the time average has to be taken, is long compared with molecular collision times, but on the other hand, not so long as those times that correspond to Poincaré cycles. With a similar procedure Kirkwood derived from the equation (3.8) of Liouville an equation of the type of the Fokker-Planck equation for $\widetilde{P}_1(\mathbf{r}_1, \mathbf{p}_1)$. A more detailed discussion of the very interesting conclusions would go beyond the scope of this survey.

Kirkwood (1947) then suggests that *the kinetic distribution function,* $f(\mathbf{r}_1, \mathbf{p}_1)$, introduced in § 2, *has to be identified with the time-smoothed average value,* $\widetilde{f}_1(\mathbf{r}_1, \mathbf{p}_1)$, according to (4.11), *of the generic distribution function,* defined by (4.2*a*),

$$\widetilde{f}_1(\mathbf{r}_1, \mathbf{p}_1, t) = N\int\int \widetilde{P}_N(\mathbf{r}^N, \mathbf{p}^N, t)\, d\mathbf{r}^{N-1}\, d\mathbf{p}^{N-1},$$

in terms of the probability density of the ensemble.

Kirkwood (1947) succeeds in deriving *Boltzmann's equation for this time-smoothed distribution function* $\widetilde{f}_1(\mathbf{p}_1)$, by integrating the equation (3.8) of Liouville for $P_N(\mathbf{r}^N, \mathbf{p}^N)$ over all coordinates and momenta except those of the molecule 1 and taking a time average of the resulting equation. Then one obtains

$$\frac{\partial \widetilde{f}_1}{\partial t} = -\mathbf{c}_1 \cdot \frac{\partial \widetilde{f}_1}{\partial \mathbf{r}_1} + \frac{1}{\tau}\int_0^\tau \int\int \frac{\partial \phi}{\partial \mathbf{r}_1} \cdot \frac{\partial f_2(\Omega_s \mathbf{r}_1, \Omega_s \mathbf{r}_2, \Omega_s \mathbf{p}_1, \Omega_s \mathbf{p}_2)}{\partial \mathbf{p}_1} d\mathbf{r}_2\, d\mathbf{p}_2\, ds. \quad \ldots\ldots (4.12)$$

Three assumptions have now to be introduced to transform this equation into Boltzmann's equation:

(i) The *binary collision hypothesis* is necessary for simplification of the integral on the right hand side. If the collisions are binary collisions in an overwhelming number of cases, as can be expected at normal gas densities, application of Liouville's equation to the collision of two molecules leads to

$$f_2(\mathbf{r}_1, \mathbf{r}_2, \mathbf{p}_1, \mathbf{p}_2, t) = f_2(\Omega_s \mathbf{r}_1, \Omega_s \mathbf{r}_2, \Omega_s \mathbf{p}_1, \Omega_s \mathbf{p}_2, t+s)$$

and also to the constancy in time of the element in phase space $d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{p}_1\, d\mathbf{p}_2$.

(ii) The *assumption about the length of the time period* τ, necessary to make the integral itself proportional to τ and the contribution to the equation, because of the factor $1/\tau$, independent of this time period. On the one hand τ must be taken long compared with molecular collision times, on the other hand τ may not be allowed to approach infinity, since then over long periods of time (Poincaré cycle periods) the orbit will be traversed in the reverse sense to any desired degree of accuracy, cancelling its initial contribution to the integral. This second assumption seems to contain essentially the irreversible element, which must necessarily come into the equation.

(iii) The *deviations from equilibrium must be small*, which allows one to replace the time average of the product of $f_1(\mathbf{p}_1)$ and $f_1(\mathbf{p}_2)$ by the product of the time averages.

After a lengthy but straightforward transformation, Kirkwood succeeds then in deriving Boltzmann's equation for the time smoothed average $\widetilde{f}_1(\mathbf{p}_1)$,

$$\frac{\partial \widetilde{f}_1}{\partial t} = -\mathbf{c}_1 \cdot \frac{\partial \widetilde{f}_1}{\partial \mathbf{r}_1} + \iint |\mathbf{c}_2 - \mathbf{c}_1| \cdot [\widetilde{f}_1(\mathbf{p}_1')\widetilde{f}_1(\mathbf{p}_2') - \widetilde{f}_1(\mathbf{p}_1)\widetilde{f}_1(\mathbf{p}_2)]\, 2\pi b\, db\, d\mathbf{p}_2. \quad \ldots\ldots (4.13)$$

Summarizing this section about the application of the method of statistical mechanics to non-equilibrium processes, it can be concluded, that it *appears to be necessary to identify the kinetic distribution function* $f(\mathbf{p})$, *upon which the kinetic theory is based, by some average of the corresponding distribution function of the ensemble of statistical mechanics*, i.e. *either the coarse grained distribution function, averaged over a finite element* $\delta\mathbf{r}\,\delta\mathbf{p}$ *of phase space or the time smoothed distribution function, averaged over a finite time period* τ.

Before closing this paragraph, some attention must be given to considerations of Born and Green (1946, 1947) of a quite different character. These authors try to avoid the use of the ensemble method of statistical mechanics, on which all considerations of this paragraph are based, in an attempt to extend the kinetic theory of § 2, so as to make the theory applicable to liquids or compressed gases. The considerations are based on the distribution function $f_N(\mathbf{r}^N, \mathbf{p}^N)$, which is by definition the probability density for the occurrence of the N molecules in any order

at the positions $\mathbf{r}^N$, $\mathbf{p}^N$ in μ-space. From their argument it is not clear whether this f_N is supposed to be a measurable quantity, as is the kinetic distribution function of §2 (in that case the experimental averaging procedure should be indicated), or a mathematical auxiliary quantity. The kinetic distribution function of §2 is identified with the integral expression

$$f_1(\mathbf{r}_1, \mathbf{p}_1) = (1/(N-1)!) \int\int f_N(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^{N-1}\, d\mathbf{p}^{N-1},$$

without any averaging procedure.

The fundamental equation for f_N

$$\frac{\partial f_N}{\partial t} = -\Sigma_i \mathbf{c}_i \cdot \frac{\partial f_N}{\partial \mathbf{r}_i} - \Sigma_i (\mathbf{X}_i + \mathbf{F}_i) \cdot \frac{\partial f_N}{\partial p_i}, \qquad \ldots\ldots (4.14)$$

which is identical with Liouville's equation, for the corresponding function $f_N(\mathbf{r}^N, \mathbf{p}^N)$ of the ensemble theory of statistical mechanics, is derived *assuming continuity of probability density in μ-space* and by using the equations of motion of the molecules (cf. §3(i)).

In the derivation of the *equilibrium solutions* for $f_N(\mathbf{r}^N, \mathbf{p}^N)$ the same difficulty appeared, as when Maxwell and Boltzmann tried to prove that the velocity distribution was a *necessary* solution, before the discovery of the method of statistical mechanics (cf. §2(iii)). As Born and Green wanted to avoid the use of statistical mechanics, they also used the H-theorem, but in a generalized form, defining $H = \int\int f_N \ln f_N\, d\mathbf{r}^N\, d\mathbf{p}^N$. The difficulty that, by using this definition of H and using equation (4.14) for $\partial f_N/\partial t$, H becomes a constant in time, was then met, somewhat artificially, by introducing a $(N+1)$th molecule, which causes H to decrease with time.

The *equation of Boltzmann* is derived by integrating equation (4.14) over the coordinates and momenta of all molecules but one, which gives

$$\frac{\partial f_1}{\partial t} = -\mathbf{c}_1 \cdot \frac{\partial f_1}{\partial \mathbf{r}_1} + \int\int \frac{\partial f_2}{\partial \mathbf{p}_1} \cdot \frac{\partial \phi(\mathbf{r}_{12})}{\partial \mathbf{r}_1}\, d\mathbf{r}_2\, d\mathbf{p}_2, \qquad \ldots\ldots (4.15)$$

and by transforming this equation, using the following three assumptions:

(i) The *hypothesis of binary collisions* (as was also done by Kirkwood). This allows one to make use of Liouville's theorem for a system of two molecules:

$$0 = (\mathbf{c}_1 - \mathbf{c}_2) \frac{\partial f_2}{\partial \mathbf{r}_{12}} + \frac{\partial \phi(\mathbf{r}_{12})}{\partial \mathbf{r}_{12}} \cdot \left(\frac{\partial f_2}{\partial \mathbf{p}_2} - \frac{\partial f_2}{\partial \mathbf{p}_1} \right),$$

in which $\partial f_2/\partial t$ is taken to be zero contrary to Kirkwood). After subtraction of the zero expression $\int\int \partial f_2/\partial \mathbf{p}_2 \cdot \partial\phi/\partial \mathbf{r}_{12}\, d\mathbf{r}_2\, d\mathbf{p}_2 = 0$ from (4.15), Liouville's equation then allows transformation of the resulting equation into

$$\frac{\partial f_1}{\partial t} = -\mathbf{c}_1 \cdot \frac{\partial f_1}{\partial \mathbf{r}_1} + \int\int (\mathbf{c}_2 - \mathbf{c}_1) \cdot \frac{\partial f_2}{\partial \mathbf{r}_{12}} \cdot d\mathbf{r}_2\, d\mathbf{p}_2. \qquad \ldots\ldots (4.16)$$

(ii) It is assumed that *in the limit* $|\mathbf{r}_{12}| = \infty$ *before and after the collision the function* $f_2(\mathbf{r}_{12}, \mathbf{p}_1, \mathbf{p}_2)$ *can be written as the product* $f_1(\mathbf{p}_1^0) f_1(\mathbf{p}_2^0)$ *and* $f_1(\mathbf{p}_1') f_1(\mathbf{p}_2')$ *respectively*. The integration can then be carried out over $\mathbf{r}_{12}$ and when $f_1(\mathbf{p}_1)$ is identified with the kinetic distribution function $f(\mathbf{p}_1)$ the resulting equation is identical with Boltzmann's equation.

Although the necessity of the first hypothesis is obvious, the bearing of the other assumptions is not clearly discussed. A derivation of Boltzmann's equation from Liouville's equation along similar lines has already been attempted by Yvon in 1935. The author of this survey believes however that for a physically rigorous derivation of Boltzmann's equation from Liouville's equation the use of some kind of averaging procedure cannot be avoided.

II. MOLECULAR DISTRIBUTION IN EQUILIBRIUM

§5. CLASSICAL THEORY

(i) *Canonical Ensemble Distribution Functions*

As has been discussed in §3 (ii) the classical probability density $P_N(\mathbf{r}^N, \mathbf{p}^N)$ of the ensemble should be a function of the constants of motion of the system if the ensemble has to represent a system in equilibrium. It has there been stated, that instead of taking the microcanonical ensemble of Boltzmann, it is of advantage to use Gibbs' *canonical ensemble*

$$P_N(\mathbf{r}^N, \mathbf{p}^N) = \frac{1}{N!\, h^{3N} Z_N} \exp\left(-\frac{H(\mathbf{r}^N, \mathbf{p}^N)}{kT}\right). \qquad \ldots\ldots (5.1)$$

The constant Z_N is a normalizing constant, T is the temperature, k Boltzmann's constant, the factor h^{-3N} has been inserted for reasons to be discussed later (§6) and the factor $1/N!$ has been inserted to ensure the additivity of entropy. As has been discussed in detail already (§3 (ii)), Gibbs' canonical ensemble has the advantage that continuous integration is possible over the entire phase space. As practically all the systems of the aggregate have an energy equal to the average energy of the system (the fluctuation $\overline{(E-\bar{E})^2}/\bar{E}^2$ being of the order of N^{-1}), the result for all calculations of averages with this canonical ensemble is equal to the average calculated with a Boltzmann microcanonical ensemble. Whereas the microcanonical ensemble should be the accurate representation of a system enclosed in a vessel with adiabatic walls, i.e. a system of absolutely constant energy, the canonical ensemble represents a system which is in temperature equilibrium with a thermostate: it can be shown that when a system, represented by a microcanonical ensemble, is divided into a small part and a large part separated from each other by heat-transmitting walls, the small part is described by a canonical ensemble and the large part acts as the thermostate.

The normalizing constant Z_N of the distribution function is determined by the normalizing condition $\int\int P_N\, d\mathbf{r}^N\, d\mathbf{p}^N = 1$ or

$$Z_N = \frac{1}{N!\, h^{3N}} \int\int \exp\left(-\frac{H(\mathbf{r}^N, \mathbf{p}^N)}{kT}\right) d\mathbf{r}^N\, d\mathbf{p}^N. \qquad \ldots\ldots (5.2)$$

The right-hand side of this equation is often called the partition function or integral over states.

In classical theory, the Hamiltonian can always be written as a sum of the kinetic energy $\Sigma \mathbf{p}_i^2/2m$ and a potential energy $\Phi(\mathbf{r}^N)$ depending on the configuration $\mathbf{r}^N \equiv (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ of all the molecules of the system. The integration over the momenta can be carried out first, giving (using (3.3)) for the configurational distribution function

$$P_N(\mathbf{r}^N) = \frac{1}{N!\, Q_N} \exp\left(-\frac{\Phi(\mathbf{r}^N)}{kT}\right), \qquad \ldots\ldots (5.3)$$

where now Q_N is the normalization constant given by

$$Q_N = \frac{1}{N!}\int \exp\left(-\frac{\Phi(\mathbf{r}^N)}{kT}\right) d\mathbf{r}^N. \qquad \ldots\ldots(5.4a)$$

The quantity Q_N is often called the configurational partition function. The relation between the two partition functions is simply

$$Q_N = \lambda^{3N} Z_N, \qquad \ldots\ldots(5.4b)$$

where $\lambda^2 = h^2/2\pi mkT$. Substitution of the equilibrium expression (5.3) for $P_N(\mathbf{r}^N)$ in equation (3.4) for $P_h(\mathbf{r}^h)$ or (3.13) for $n_h(\mathbf{r}^h)$ gives the following general expression for the configurational distribution function of a group of h molecules:

$$n_h(\mathbf{r}^h) = \frac{N!}{(N-h)!}\int P_N(\mathbf{r}^N)\, d\mathbf{r}^{N-h} = \frac{1}{(N-h)!\, Q_N}\int \ldots\ldots \int \exp -\frac{\Phi(\mathbf{r}^N)}{kT}\, d\mathbf{r}^{N-h}. \qquad \ldots\ldots(5.5)$$

It will be shown in the next sub-section (ii) that the thermal and caloric equation of state can both be expressed in terms of this distribution function for the special case of $h=2$, which will be called the *molecular pair distribution function.* In the next section (§ 6) it will be shown how this equation has to be written in quantum statistical mechanics.

(ii) *Caloric and Thermal Equation of State*

Using the Gibbs' canonical ensemble, the average value of some quantity G which is a function of the values of the momenta and coordinates of all molecules $G(\mathbf{r}^N, \mathbf{p}^N)$, is given by the equation

$$\bar{G} = \frac{1}{N!h^{3N}Z_N}\int\int \exp\left(-\frac{H(\mathbf{r}^N, \mathbf{p}^N)}{kT}\right) G(\mathbf{r}^N, \mathbf{p}^N)\, d\mathbf{r}^N\, d\mathbf{p}^N. \qquad \ldots\ldots(5.6)$$

Two important averages, which dominate the field of the thermal and caloric properties of fluids, are the average values obtained for the energy and the pV-value. The kinetic theory states that the energy of the system is equal to the average value of the Hamiltonian $\Sigma \mathbf{p}_i^2/2m + \Phi(\mathbf{r}^N)$. Assuming that no external field is present and that the potential energy can be written as a double sum over the potential energy of all pairs of molecules $\phi(r_{i\kappa})$, in which $r_{i\kappa} = |\mathbf{r}_{i\kappa}| = |\mathbf{r}_\kappa - \mathbf{r}_i|$, we obtain

$$U = \frac{1}{N!\, h^{3N} Z_N}\int\int \exp\left(-\frac{H}{kT}\right)\left(\Sigma_i \frac{p_i^2}{2m} + \tfrac{1}{2}\Sigma_i\Sigma_\kappa \phi(r_{i\kappa})\right) d\mathbf{r}^N\, d\mathbf{p}^N. \qquad \ldots\ldots(5.7)$$

The integration over the momenta can be carried out giving for the first term $3NkT/2 = 3RT/2$. Integration of the second term over the coordinates of all molecules except the molecules i and κ for each term $\phi(r_{i\kappa})$ leads to $N(N-1)$ similar terms $\frac{1}{2}\int\int P_2(\mathbf{r}_i, \mathbf{r}_\kappa)\phi(r_{i\kappa})\, d\mathbf{r}_i\, d\mathbf{r}_\kappa$. Introduction of the generic distribution functions $n_2(\mathbf{r}_i, \mathbf{r}_\kappa) = N(N-1)P_2(\mathbf{r}_i, \mathbf{r}_\kappa)$ finally gives

$$U = \tfrac{3}{2}RT + \tfrac{1}{2}\int\int n_2(\mathbf{r}_1, \mathbf{r}_2)\phi(r_{12})\, d\mathbf{r}_1\, d\mathbf{r}_2. \qquad \ldots\ldots(5.8)$$

This equation gives U as a function of temperature and volume, since $n_2(\mathbf{r}_1, \mathbf{r}_2)$ is a function of V and T and is called *the caloric equation of state.*

The *virial theorem* gives a kinetic expression for the pressure-volume product (pV-value) of the fluid. This virial theorem states that for a system in equilibrium the average value of the virial of the external and internal forces exerted on the

molecular system is equal to minus twice the average value of the kinetic energy of the system. If this theorem is applied to a fluid enclosed in a vessel of volume V under an external pressure p the virial of the external forces is $-3pV$, the virial of the intermolecular forces Ξ is equal to $\Xi(\mathbf{r}^N) = \Sigma \mathbf{r}_i \,.\, \mathbf{F}_i = -\frac{1}{2}\Sigma\Sigma r_{i\varkappa} \,.\, d\phi/dr_{i\varkappa}$. The kinetic energy K is $\Sigma p_i^2/2m$. So the virial theorem states that $-2\overline{K} = -3pV + \overline{\Xi}$ or

$$2\Sigma_i \overline{\frac{p_i^2}{2m}} = 3pV + \tfrac{1}{2}\Sigma_i \Sigma_\varkappa r_{i\varkappa} \frac{\partial \phi}{dr_{i\varkappa}} \,. \qquad \ldots\ldots(5.9)$$

Solving this equation for pV and expressing the average in terms of the canonical aggregate, we obtain

$$pV = \frac{1}{N!\,h^{3N} Z_N} \int\int \exp\left(-\frac{H}{kT}\right) \cdot \left(\tfrac{2}{3}\Sigma_i \frac{p_i^2}{2m} - \tfrac{1}{3}\Sigma_i \Sigma_\varkappa r_{i\varkappa} \frac{\partial \phi}{dr_{i\varkappa}}\right) d\mathbf{r}^N \, d\mathbf{p}^N. \qquad \ldots\ldots(5.10)$$

Carrying out the integration, one obtains in similar manner to the transformation of (5.7) into (5.8) the expression

$$pV = RT - \tfrac{1}{3} \int\int n_2(\mathbf{r}_1, \mathbf{r}_2) r_{i\varkappa} \frac{\partial \phi}{dr_{i\varkappa}} \, d\mathbf{r}_1 \, d\mathbf{r}_2. \qquad \ldots\ldots(5.11)$$

This equation, giving the value of pV as function of V and T, will be called the *thermal equation of state*.

Both the caloric and the thermal equation of state are expressed as the sum of a term caused by the kinetic energy and a contribution of the molecular interaction, expressed in terms of the distribution function for a pair of molecules. This shows the importance of knowing the properties of this distribution function. It is given in terms of the distribution function of N molecules by the formula (5.5):

$$n_2(\mathbf{r}_1, \mathbf{r}_2) = \frac{N(N-1)}{N!\,Q_N} \int \exp\left(-\frac{\Phi(\mathbf{r}^N)}{kT}\right) d\mathbf{r}_3 \ldots . \, d\mathbf{r}_N. \qquad \ldots\ldots(5.12)$$

It will be shown in the next section that the expressions (5.11) and (5.8) for the thermal and caloric equation of state can also be derived with the help of statistical thermodynamics.

(iii) *Statistical Thermodynamics with the Canonical Ensemble*

The relationship between this method of statistical mechanics and the thermodynamic functions is formed by the statement that the thermodynamic function of Helmholtz, the free energy $F = U - TS$ is equal to $-kT \ln Z_N$, where Z_N is the normalizing constant of the distribution function P_N. (*Thermodynamic analogy of Gibbs*). (The formulae are still more elegant if the Planck function $\Psi = S - U/T$ is introduced, which is equal to $-F/T$.) The fundamental relation between thermodynamics and statistical mechanics is the relation

$$F = -kT \ln Z_N, \qquad \ldots\ldots(5.13)$$

where Z_N is given by (5.2). The basic thermodynamical expression for the Helmholtz function is the differential expression

$$dF = -S\,dT - p\,dV - \mu\,dN, \qquad \ldots\ldots(5.14)$$

giving, with $U = F + TS$,

$$U = \left(\frac{\partial(F/T)}{\partial(1/T)}\right)_{V,N}, \quad p = -\left(\frac{\partial F}{\partial V}\right)_{T,N}, \quad \mu = \left(\frac{\partial F}{\partial N}\right)_{V,T}. \quad \ldots\ldots(5.15)$$

This leads also to expressions for the energy and the pressure which should be compared with the previous ones.

The energy becomes, substituting the partition function (5.2) for Z_N,

$$U = -k\frac{\partial \ln Z_N}{\partial(1/T)} = \frac{1}{N!\,h^{3N}Z_N}\int\int \exp\left(-\frac{H}{kT}\right).H\,d\mathbf{r}^N\,d\mathbf{p}^N, \quad \ldots\ldots(5.16)$$

which expression is identical with the equation (5.7) for the energy derived above.

The second expression for the pressure becomes $p = kT\partial \ln Z_N/\partial V$. Green (1947) has shown how to transform the partition function so that the differentiation with respect to V can be carried out. The difficulty that the volume V occurs in the limits of the partition function can be overcome by transforming to new space coordinates $\bar{\mathbf{r}} = \mathbf{r}/L$, where L is a scale factor (a characteristic length of the vessel, the volume is proportional to L^3). The partition integral then transforms into $L^{3N}\int\int \exp -\beta H)(L\bar{\mathbf{r}}^N, \mathbf{p}^N)\,d\bar{\mathbf{r}}^N\,d\mathbf{p}^N$, where V or L no longer occurs in the limits of the integral. Then

$$pV = kT\frac{1}{3}\frac{L}{Z_N}\frac{\partial Z_N}{\partial L} = \frac{kT^{3N}}{N!\;h^{3N}Z_N}\int\int \exp\left(-\frac{H}{kT}\right)\cdot\left(N - \frac{1}{kT}\frac{L}{3}\frac{\partial H}{\partial L}\right)d\bar{\mathbf{r}}^N\,d\mathbf{p}^N. \quad \ldots\ldots(5.17)$$

The first term in the integrand gives, transforming back to $\mathbf{r}^N$ as coordinates, kTN. The second term can be transformed, remembering that L occurs in H only in the combination $L\bar{\mathbf{r}}_{i\alpha} = \mathbf{r}_{i\alpha}$, which occurs in $\phi(\mathbf{r}_{i\alpha})$. This makes the differential quotient $\frac{1}{3}L\,\partial H/\partial L = \frac{1}{6}\Sigma\Sigma L\,\partial\phi_{i\alpha}/\partial L = \frac{1}{6}\Sigma\Sigma r_{i\alpha}\,d\phi/dr_{i\alpha}$. Thus the equation (5.17) can be transformed into

$$pV = RT - \frac{1}{6}\frac{1}{N!\;h^{3N}Z_N}\int\int \exp\left(-\frac{H}{kT}\right)\Sigma_i\Sigma_\kappa \mathbf{r}_{i\kappa}\frac{\partial\phi}{dr_{i\kappa}}\,d\mathbf{r}^N\,d\mathbf{p}^N, \quad \ldots\ldots(5.18)$$

which is identical with (5.10). Thus both the thermal and the caloric equations of state are proved to be the same in both methods. *This can, on the other hand, be considered as a proof that* $T\ln Z$ *is indeed proportional to Helmholtz' function* F.

The third equation gives the thermodynamic potential μ in terms of the partition function; remembering that N is a very large number this equation becomes

$$\mu = F_N - F_{N-1} = kT\ln(Z_{N-1}/Z_N). \quad \ldots\ldots(5.19)$$

Often, however, the *activity* z is used instead of the thermodynamic potential μ, being defined as

$$z = Q_{N-1}/Q_N. \quad \ldots\ldots(5.20)$$

The relation between μ and z can be obtained from (5.4*b*) to be given by the equation

$$\mu = kT\ln z + \tfrac{3}{2}kT\ln(h^2/2\pi mkT). \quad \ldots\ldots(5.21)$$

The activity so defined has according to (5.21) the dimension number of molecules per unit volume. For an ideal gas for instance $\Phi = 0$, $Q_N = V^N/N!$ and $z = N/V = n$. For compressed gases z approaches this limit n only for large dilution. Theoretical formulae become considerably simpler when the activity z is introduced as an independent variable instead of the molecular number density n of the fluid.

Finally the expression (5.1) for $P_N(\mathbf{r}^N, \mathbf{p}^N)$ is sufficient to evaluate also the statistical mechanical expression (4.3) for the H-function of Boltzmann for a system in equilibrium. This becomes, using (5.13), (5.16) and $S = -F/T + U/T$,

$$H = -S/k - \ln(N!\, h^{3N}), \qquad \ldots\ldots(5.22)$$

which proves that the H-function is equal to minus the entropy of the system divided by the Boltzmann constant k, except for an unimportant numerical additive constant.

§6. QUANTUM THEORY

(i) *Canonical Ensemble*

In quantum statistical mechanics the probability density of the ensemble has to be replaced by the statistical operator $\mathscr{P}$, according to § 3 (iv). For a system in equilibrium this operator $\mathscr{P}$ is chosen to be

$$\mathscr{P} = \frac{1}{Z_N} \exp\left(-\frac{\mathscr{H}(p^N, \mathbf{r}^N)}{kT}\right), \qquad \ldots\ldots(6.1)$$

where $\mathscr{H}$ is the Hamilton operator, obtained by replacing in the Hamilton function all momenta $\mathbf{p}_i$ by the corresponding operators $p_i = -i\hbar\partial/\partial\mathbf{r}_i$. The operator $\exp -\beta\mathscr{H}$ is defined to be $\Sigma - (\beta\mathscr{H})^n/n!$ and Z_N is again the normalizing factor.

According to this the *density matrix* (3.20) becomes

$$\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^{N\prime}) = \frac{1}{Z_N} \Sigma_\varrho \phi_\varrho^*(\mathbf{r}^{N\prime}) \exp\left(-\frac{\mathscr{H}}{kT}\right) \phi_\varrho(\mathbf{r}^N) = \frac{1}{Z_N} \exp\left(-\frac{\mathscr{H}}{kT}\right)(\mathbf{r}^{N\prime}, \mathbf{r}^{N\prime}). \qquad \ldots\ldots(6.2)$$

The diagonal elements of this matrix, obtained by putting $\mathbf{r}^{N\prime} = \mathbf{r}^N$, are

$$\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^N) = \frac{1}{Z_N} \Sigma_\varrho \phi_\varrho^* \exp\left(-\frac{\mathscr{H}}{kT}\right) \phi_\varrho. \qquad \ldots\ldots(6.3)$$

This expression is the *quantum mechanical expression for the probability density for a configuration* $\mathbf{r}^N$, the classical expression of which is given by (5.3). The important property of (6.2) as well as (6.3) is the fact that they are *independent of the particular system of orthonormal wave functions* ϕ_ϱ, chosen to describe them (cf. § 3 (iv), (3.19)). If in particular energy eigenfunctions ψ_σ are chosen $\mathscr{H}\psi_\sigma = E_\sigma\psi_\sigma$, the density matrix becomes a diagonal matrix, with matrix elements

$$\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^N) = \frac{1}{Z_N} \Sigma_\sigma |\psi_\sigma(\mathbf{r}^N)|^2 \exp\left(-\frac{E_\sigma}{kT}\right). \qquad \ldots\ldots(6.4)$$

Here $\exp(-E_\sigma k/T)$ represents the probability that a state σ occurs, and $|\psi_\sigma|^2$ denotes the probability density that in this state σ the system has the configuration $\mathbf{r}^N$, which shows that the expression is really the quantum mechanical configurational probability density. In sub-section (iii) it will be shown how expression (6.4) approaches the corresponding classical expression $P_N(\mathbf{r}^N)$ given by (5.3) at high temperatures, where deviations from quantum theory are small.

We also introduce in § 3 the probability density matrix $\mathscr{P}_{\varrho\varrho'}$ which now takes the form

$$\mathscr{P}_{\varrho\varrho'} = \overline{c_\varrho^* c_{\varrho'}} = \frac{1}{Z_N} \int \phi_\varrho^* \exp\left(-\frac{\mathscr{H}}{kT}\right) \phi_{\varrho'}\, d\mathbf{r}^N. \qquad \ldots\ldots(6.5)$$

This *matrix depends on the system of orthonormal functions* which is chosen as a base: change to another system, $\chi_\tau = \Sigma \phi_\varrho \alpha_{\varrho\tau}$, gives $\mathscr{P}_{\tau\tau'} = \Sigma\Sigma \alpha^{-1}_{\tau\varrho} \mathscr{P}_{\varrho\varrho'} \alpha_{\varrho'\tau'}$ in agreement with the general rules of matrix transformation theory. In the language of energy wave functions $\mathscr{P}_{\sigma\sigma'}$ takes the diagonal form

$$\mathscr{P}_{\sigma\sigma'} = \delta_{\sigma\sigma'} \frac{1}{Z_N} \exp\left(-\frac{E_\sigma}{kT}\right). \qquad \ldots\ldots(6.6)$$

Here $\mathscr{P}_{\sigma\sigma}$ is the *probability that the system is found to be in the eigenstate σ with the energy E_σ*.

The expression (6.3) for the probability density for a configuration $\mathbf{r}^N$ and the expression (6.6) for the probability for an eigenstate σ can be normalized by taking the trace of the matrix and equalizing this to unity. Taking the trace means summing $\mathscr{P}_{\sigma\sigma}$ over σ in (6.6) and integrating $\mathscr{P}(\mathbf{r}^N \quad \mathbf{r}^N)$ over $\mathbf{r}^N$ in (6.3). This trace is invariant for the particular representation chosen. Thus

$$\Sigma \mathscr{P}_{\sigma\sigma} = \int \mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^N)\, d\mathbf{r}^N = 1,$$

giving

$$Z_N = \int \left(\Sigma_\varrho \phi^*_\varrho(\mathbf{r}^N) \exp\left(-\frac{\mathscr{H}}{kT}\right) \phi_\varrho(\mathbf{r}^N)\right) d\mathbf{r}^N = \Sigma_\sigma \exp\left(-\frac{E_\sigma}{kT}\right). \qquad \ldots\ldots(6.7)$$

This is the quantum mechanical expression of the partition function which here takes the place of the expression (5.2) of classical theory. According to the second form it is often called the sum over states, but then too much weight is laid on the energy eigenfunctions used. The first form shows that the partition function can also be written as an integral over configuration space alone.

(ii) *Caloric and Thermal Equation of State*

According to § 3 (v) the average value of some quantity G, which, being given in terms of $\mathbf{r}^N$ and $\mathbf{p}^N$ by $G(\mathbf{r}^N, \mathbf{p}^N)$, is represented in quantum theory by the operator function $\mathscr{G}(\mathbf{r}^N, p^N)$, is given by the formula:

$$\overline{G} = \int (\mathscr{G}\mathscr{P})\,(\mathbf{r}^N;\ \mathbf{r}^N)\, d\mathbf{r}^N = \frac{1}{Z_N} \int \left(\Sigma_\varrho \phi^*_\varrho\, \mathscr{G} \exp\left(-\frac{\mathscr{H}}{kT}\right) \phi_\varrho\right) d\mathbf{r}^N, \qquad \ldots\ldots(6.8)$$

or by the expressions

$$\overline{G} = \Sigma_\varrho (\mathscr{G}\mathscr{P})_{\varrho\varrho} = \frac{1}{Z_N} \Sigma G_{\sigma\sigma} \exp\left(-\frac{E_\sigma}{kT}\right). \qquad \ldots\ldots(6.9)$$

Again, equations (6.8) and (6.9) show that $\overline{G}$, being the trace either of the matrix $(\mathscr{G}\mathscr{P})(\mathbf{r}^N;\ \mathbf{r}^{N'})$ or of the matrix $(\mathscr{G}\mathscr{P})_{\varrho\varrho'}$, is independent of the system of orthonormal wave functions which is chosen as a basis for the calculation of the matrix elements as it of course should be.

The *caloric equation of state* is, according to (6.8), given by

$$U = \frac{1}{Z_N} \int \left((\mathscr{H} + \Phi) \exp\left(-\frac{\mathscr{H}}{kT}\right)\right)(\mathbf{r}^N;\ \mathbf{r}^N)\, d\mathbf{r}^N. \qquad \ldots\ldots(6.10)$$

The derivation of the thermal equation of state has to be based on the virial theorem. In quantum theory this theorem has been proved to be valid also by Slater (1933) provided the expectation values for the kinetic energy and for the virial of the internal and external forces are introduced, which of course has to be done in all quantum mechanical calculations.

With the virial theorem, $3pV$ was shown to be equal to the sum of the average value of twice the kinetic energy and the average value of the virial expression for the intermolecular forces (eqn. 5.9).

Using the method of calculating average values with the density matrix given by (6.8) the *thermal equation of state* now becomes

$$pV = \frac{1}{Z_N}\int\left(\left(\tfrac{2}{3}\mathcal{K} + \tfrac{1}{3}\Xi\right)\exp\left(-\frac{\mathcal{H}}{kT}\right)\right)(\mathbf{r}^N;\ \mathbf{r}^N)\,d\mathbf{r}^N. \quad \ldots\ldots(6.11)$$

Again, it can be shown that these expressions for U and pV can be transformed into the expressions obtained from statistical thermodynamics, starting from the fundamental relationship that Helmholtz' function $F = -kT\ln Z_N$. This is important in connection with the doubt that has been raised by Born and Green (1947 c) against this equality of the "thermodynamic" and "kinetic" pressure in quantum theory and quantum hydrodynamics and which they have used as a tool to explain the strange phenomena in liquid helium (Born and Green 1948).

According to the formulae of statistical thermodynamics (5.15), the energy and the pressure should be equal respectively to

$$U = -\frac{k}{Z_N}\frac{\partial}{\partial(1/T)}\int\exp\left(-\frac{\mathcal{H}}{kT}\right)(\mathbf{r}^N;\ \mathbf{r}^N)\,d\mathbf{r}^N, \quad \ldots\ldots(6.12)$$

$$PV = -\frac{kT}{Z_N}V\frac{\partial}{\partial V}\int\exp\left(-\frac{\mathcal{H}}{kT}\right)(\mathbf{r}^N;\ \mathbf{r}^N)\,d\mathbf{r}^N. \quad \ldots\ldots(6.13)$$

The first of these two equations is identical with (6.10), as can be directly verified by differentiating $\exp(-\mathcal{H}/kT)$ with respect to $-1/kT$. The identity of (6.13) with (6.11) has to be proved as in classical theory by introducing a scale factor. One must take care, however, to investigate how the density matrix transforms. This can be done most easily by substituting a particular system of wave functions,

$$\left(\exp-\frac{\mathcal{H}}{kT}\right)(\mathbf{r}^N;\ \mathbf{r}^N) = \int\ldots\int d\mathbf{k}_1\ldots d\mathbf{k}_N\exp(\Sigma i\mathbf{k}_\iota.\mathbf{r}_\iota)\exp\left(-\frac{\mathcal{H}}{kT}\right)\exp(\Sigma - i\mathbf{k}_\iota.\mathbf{r}_\iota). \quad \ldots\ldots(6.14)$$

This is allowed because in the density matrix *every* orthonormal and complete system of eigenfunctions may be substituted.

Transforming now to reduced coordinates $\mathbf{r} = L\bar{\mathbf{r}}$, etc. one has to write $\mathbf{k} = L^{-1}\bar{\mathbf{k}}$ and accordingly:

$$pV = -\frac{1}{Z_N}V\frac{\partial}{\partial V}\int\ldots\int\exp(\Sigma i\bar{\mathbf{k}}_\iota.\bar{\mathbf{r}}_\iota)\exp\left(-\frac{\mathcal{H}}{kT}\right) \times\exp(-\Sigma i\bar{\mathbf{k}}_\iota.\bar{\mathbf{r}}_\iota)\,d\bar{\mathbf{k}}_1\ldots d\bar{\mathbf{k}}_N\,d\bar{\mathbf{r}}_1\ldots d\bar{\mathbf{r}}_N. \quad \ldots\ldots(6.15)$$

Thus the scale factor L occurs only in the Hamilton operator

$$\mathcal{H} = -\Sigma_\iota\frac{\hbar^2}{2mL^2}\bar{\Delta} + \tfrac{1}{2}\Sigma_\iota\Sigma_\kappa\phi(\bar{r}_{\iota\kappa}L).$$

Accordingly

$$V\frac{\partial\mathcal{H}}{\partial V} = \tfrac{1}{3}L\frac{\partial\mathcal{H}}{\partial L} = -\tfrac{2}{3}\mathcal{K} + \tfrac{1}{6}\Sigma_\iota\Sigma_\kappa r_{\iota\kappa}\frac{d\phi(r_{\iota\kappa})}{dr_{\iota\kappa}} = -\tfrac{2}{3}\mathcal{K} - \tfrac{1}{6}\Xi, \quad \ldots\ldots(6.16)$$

carrying out the differentiation in the exponential and remembering that the order of the operators in the integral expression can be interchanged, one obtains easily

equation (6.11) (cf. de Boer 1949a)). So both in quantum statistical mechanics and in classical statistical mechanics the expressions obtained for the caloric and thermal equations of state are exactly the same, whether one starts from the statistical expressions or makes use of the thermodynamical partition function method.

The two equations can be simplified, using reduced density matrices n_1 and n_2, defined in (3.25)

$$U = \frac{1}{2m}\int (p_1^2 n_1)(\mathbf{r}_1;\ \mathbf{r}_1)\, d\mathbf{r}_1 + \tfrac{1}{2}\int \phi(r_{12}) n_2(\mathbf{r}_1\mathbf{r}_2;\ \mathbf{r}_1\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2, \qquad \ldots\ldots(6.17)$$

$$pV = \frac{1}{3m}\int (p_1^2 n_1)(\mathbf{r}_1;\ \mathbf{r}_1)\, d\mathbf{r}_1 - \tfrac{1}{6}\int r_{12}\cdot\frac{\partial\phi}{\partial r_{12}} n_2(\mathbf{r}_1\mathbf{r}_2;\ \mathbf{r}_1\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2. \qquad \ldots\ldots(6.18)$$

When these two expressions are compared with the corresponding classical expression (5.8) and (5.11) the following differences can be remarked: (i) the first term $3RT/2$ or RT is replaced in quantum theory by the quantum mechanical average value of the kinetic energy and $\frac{2}{3}$ of this value respectively; (*b*) the second term differs only from the classical one in that the classical pair distribution function is replaced by its quantum mechanical analogue, the diagonal elements of the density matrix corresponding to a pair of molecules.

(iii) *Correspondence between Classical and Quantum Theory at High Temperatures*

According to (5.1), the *classical* expression for the probability density of a configuration $\mathbf{r}^N$ is given by

$$P_N(\mathbf{r}^N) = \frac{1}{\lambda^{3N} N!\, Z_N}\exp\left(-\frac{\Phi(\mathbf{r}^N)}{kT}\right), \qquad \ldots\ldots(6.19)$$

where expression (5.4*b*) is used for Q_N and where λ^2 is defined to be $h^2/2\pi mkT$. The *quantum mechanical analogue* of this expression should be, according to §3, the diagonal elements of the quantum mechanical density matrix (6.3)

$$\mathscr{P}_N(\mathbf{r}^N;\ \mathbf{r}^N) = \frac{1}{Z_N}\Sigma_\varrho \phi_\varrho^*(\mathbf{r}^N)\exp\left(-\frac{\mathscr{H}}{kT}\right)\phi_\varrho(\mathbf{r}^N). \qquad \ldots\ldots(6.20)$$

It is of advantage to write this in the form

$$\mathscr{P}(\mathbf{r}^N;\ \mathbf{r}^N) = \frac{1}{\lambda^{3N} N!\, Z_N}\, W(\mathbf{r}^N), \qquad \ldots\ldots(6.21)$$

because then the "Slater sum", named after Slater (1931), given by

$$W(\mathbf{r}^N) = N!\,\lambda^{3N}\Sigma_\varrho \phi_\varrho^*(\mathbf{r}^N)\exp\left(-\frac{\mathscr{H}}{kT}\right)\phi_\varrho(\mathbf{r}^N) \qquad \ldots\ldots(6.22)$$

is the exact quantum mechanical analogue of the function $\exp(-\Phi/kT)$: it can be shown that $W(\mathbf{r}^N)$ and $\exp(-\Phi(\mathbf{r}^N)/kT)$ approach each other in the region of high temperatures, where the deviations due to the quantum theory are small and unimportant (cf. Kirkwood 1931, Wigner 1932, Uhlenbeck 1935, Uhlenbeck and Gropper 1932, Husimi 1940).

This proof can be given by using the important property of expression (6.22) of being *invariant* for the system of orthonormal eigenfunctions ϕ_ϱ used. Kirkwood (1933) substituted the exponential orthonormal set of functions

$$\psi_\sigma = \frac{1}{(V^N N!)^{\frac{1}{2}}}\Sigma_{\mathbf{p}}\delta_{\mathbf{p}}\exp 2\pi i\boldsymbol{\sigma}^N.\mathbf{P}\mathbf{r}^N. \qquad \ldots\ldots(6.23)$$

The eigenfunction has been made symmetrical or antisymmetrical in the indices of the N particles of the system by summation over all permutations P of the indices. $\mathrm{P}\mathbf{r}^N$ indicates the $3N$ dimensional vector which results from the application of this permutation of indices to the vector $\mathbf{r}^N$. δ is equal to $+1$ for the symmetrical case (Bose-Einstein statistics) and $(-1)^p$ for the antisymmetrical case (Fermi-Dirac statistics), where p is the order of the permutation P. The factor $V^{-N/2}$ before the expression results from the normalization of the eigenfunction in the volume $V=L^3$. The wave vector $\boldsymbol{\sigma}^N$, having components $\sigma_{x1}, \sigma_{y1}, \sigma_{z1}, \ldots,$ $\sigma_{xN}, \sigma_{yN}, \sigma_{zN}$, takes the values n/L for each of the components, where n is an integer. The running waves correspond to a stream of particles with momentum $\mathbf{p}=\hbar\boldsymbol{\sigma}$.

Substitution of this expression (6.23) in (6.22) for the eigenfunctions ϕ_ϱ gives, replacing the summation over $\boldsymbol{\sigma}$ by an integration over the momentum $\mathbf{p}$:

$$W(\mathbf{r}^N)=\left(\frac{\lambda}{\hbar}\right)^{3N}\int \Sigma_{\mathrm{P}}\delta_{\mathrm{P}}\exp\left(-\frac{i\mathbf{p}^N\,.\,\mathrm{P}\mathbf{r}^N}{\hbar}\right)\exp\left(-\frac{\mathscr{H}}{kT}\right)\exp\left(\frac{i\mathbf{p}^N\,.\,\mathrm{P}\mathbf{r}^N}{\hbar}\right)d\mathbf{p}^N. \qquad \ldots\ldots(6.24)$$

In general the Hamiltonian $\mathscr{H}$ is equal to $\Sigma p_i^2/2m+\Phi(\mathbf{r}^N)$.

Consider first the case of an *ideal gas* for which $\mathscr{H}=\Sigma p_i^2/2m$. In that case the exponential functions are eigenfunctions and the operator can be replaced by the eigenvalue $\Sigma\mathbf{p}_i^2/2m$. The integrals can then be evaluated, leading to

$$W_{id}(\mathbf{r}^N)=\Sigma_{\mathrm{P}}\delta_{\mathrm{P}}\exp\left(-\frac{\pi}{\lambda^2}\,|\,\mathbf{r}^N-\mathrm{P}\mathbf{r}^N\,|^2\right). \qquad \ldots\ldots(6.25)$$

The identity amongst the permutations gives unity. So if the symmetry of the eigenfunctions has no influence $W(\mathbf{r}^N)=1$. This is the value which $W(\mathbf{r}^N)$ would have in classical theory for an ideal gas ($\Phi=0$).

Deviations from unity result from the *symmetry* of the wave functions. In the case of symmetrical wave functions the probability of finding two molecules ι and κ at a distance $r_{\iota\kappa}$ rises above the value unity when $r_{\iota\kappa}$ decreases. This will be discussed in more detail in §9(ii) and §8(iii). If attractive forces were acting between the molecules the result would be exactly the same. In the case of an ideal gas intermolecular forces are lacking and so one speaks often of the "*apparent attraction*" *between molecules with symmetrical wave functions.* The statistics followed by an ideal gas or a gas of nearly independent molecules having symmetrical wave functions are the statistics introduced by Bose and Einstein.

If the eigenfunctions are made *antisymmetrical* in the indices of the particles, $\delta_{\mathrm{P}}=(-1)^p$. The result appears to be that now the Slater sum $W(\mathbf{r}^N)$ *de*creases below the value unity, when the distance $r_{\iota\kappa}$ between two molecules ι and κ decreases, reaching the value zero when the distance is zero. Thus the probability density of finding two molecules in exactly the same place is zero, although in this case of the ideal gas no forces whatever are acting between the molecules. This effect is often indicated as an "*apparent repulsion*" *between the molecules of the gas having antisymmetrical wave functions* (which satisfy the Pauli principle). The statistics for independent molecules in this case are the so called Fermi-Dirac statistics. For more details and examples the reader is referred to Uhlenbeck and Gropper (1932) and Uhlenbeck (1935).

In both cases the distance over which the molecules influence each other due to the symmetry of wave functions is of the order of magnitude $\lambda = h/(2\pi mkT)^{\frac{1}{2}}$, which is, except for a numerical factor, the de Broglie wavelength of the molecular motion at a temperature T. As this de Broglie wavelength becomes much smaller than the molecular diameter when T increases, these deviations "due to the statistics" are quite small at high temperatures. Consequently for an ideal gas the quantum mechanical expression $W(\mathbf{r}^N)$ approaches at high temperatures to the classical expression $\exp -\Phi/kT = 1$ for an ideal gas.

For a *real gas* the complete expression for $\mathscr{H}$ has to be substituted in equation (6.24) for $W(\mathbf{r}^N)$. That at high temperatures $-W(\mathbf{r}^N)$ now also approaches the classical expression $\exp -\Phi/kT$ has been shown by Kirkwood (1931). For simplicity we neglect symmetry effects. The factor

$$X = \exp\left(-\frac{\mathscr{H}}{kT}\right) \exp\left(\frac{i}{\hbar}\,\mathbf{p}^N.\,\mathbf{r}^N\right), \qquad \ldots\ldots (6.26)$$

of the integral expression (6.24) for the Slater sum satisfies the so-called Bloch equation (1932) $\mathscr{H}X = -\partial X/\partial\beta$, where $\beta = 1/kT$; this equation must be solved with the boundary conditions $\beta = 0$, $X \to \exp(i/\hbar)\mathbf{p}^N.\,\mathbf{r}^N$. As a solution for X is substituted

$$X = \exp\left(\frac{i}{\hbar}\,\mathbf{p}^N.\,\mathbf{r}^N\right) \exp\left(-\frac{H(\mathbf{p}^N, \mathbf{r}^N)}{kT}\right) Y(\mathbf{p}^N, \mathbf{r}^N, T). \qquad \ldots\ldots (6.27)$$

The Bloch equation is then transformed into an equation for Y of the form

$$\frac{\partial Y}{\partial\beta} = \frac{i\hbar}{m}\left(\mathbf{p}^N \cdot \frac{\partial Y}{\partial \mathbf{r}^N} - \beta Y \mathbf{p}^N \cdot \frac{\partial \Phi}{\partial \mathbf{r}^N}\right) + \frac{\hbar^2}{2m}\left\{\frac{\partial}{\partial \mathbf{r}^N} \cdot \frac{\partial Y}{\partial \mathbf{r}^N} - 2\beta \frac{\partial \Phi}{\partial \mathbf{r}^N} \cdot \frac{\partial \Phi}{\partial \mathbf{r}^N}\right.$$
$$\left. - \beta Y \frac{\partial}{\partial \mathbf{r}^N} \cdot \frac{\partial \Phi}{\partial \mathbf{r}^N} + \beta^2 Y \left(\frac{\partial \Phi}{\partial \mathbf{r}^N}\right)^2\right\}. \qquad \ldots\ldots (6.28)$$

Evidently the boundary condition for Y leads to the condition that Y approaches unity for $\beta \to 0$. This equation for $Y(\mathbf{p}^N, \mathbf{r}^N, T)$ can now be solved by substituting a series development for $Y = \Sigma h^n y_n(\mathbf{p}^N, \mathbf{r}^N, T)$ in powers of h and equating on both sides terms containing equal powers of h. The result must be substituted in (6.24) and has to be integrated over $\mathbf{p}^N$, which makes the terms containing odd powers of h disappear. The resulting expression for $W(\mathbf{r}^N)$ is, neglecting symmetry effects in order to avoid complication,

$$W(\mathbf{r}^N) = \exp\left(-\frac{\Phi(\mathbf{r}^N)}{kT}\right)\left\{1 + \frac{\hbar^2}{mkT} W_2(r^N, T) + \frac{\hbar^4}{m^2k^2T^2} W_4(\mathbf{r}^N, T) + \ldots.\right\}, \qquad \ldots\ldots (6.29)$$

where
$$W_2(\mathbf{r}^N, T) = \frac{1}{48\pi^2}\left(\frac{\partial}{\partial \mathbf{r}^N} \cdot \frac{\partial \Phi}{\partial \mathbf{r}^N} - \frac{1}{2kT} \frac{\partial \Phi}{\partial \mathbf{r}^N} \cdot \frac{\partial \Phi}{\partial \mathbf{r}^N}\right). \qquad \ldots\ldots (6.30)$$

and similar expressions for W_4, etc.

Thus *at high temperatures this quantum mechanical expression* $W(\mathbf{r}^N)$ *approaches* $\exp -\Phi/kT$. The series expansion (6.29) converges when the factor h^2/mkT, which is of the order of magnitude of the square of the de Broglie wavelength at a temperature T, is small. So this expression offers a good approximation for the quantum mechanical expression for $W(\mathbf{r}^N)$, when the temperature is high and will be used for this purpose in §§ 8 and 9 to evaluate the distribution function and the equation of state of gases, especially of hydrogen and deuterium, in the high temperature region, where the deviations from classical theory can be expected to be small.

The Slater sum $W_N(\mathbf{r}^N)$ introduced in this section, which has been shown to be the real quantum mechanical equivalent of the function $\exp -\Phi(\mathbf{r}^N)/kT$ occurring in the classical formulae, can be used to write the quantum mechanical formulae in closer agreement with the classical expressions. In particular, the configurational probability density $\mathscr{P}_N(\mathbf{r}^N;\ \mathbf{r}^N)$ can, according to (6.21), be written as

$$\mathscr{P}_N(\mathbf{r}^N;\ \mathbf{r}^N) = \frac{1}{N!\ Q_N} W_N(\mathbf{r}^N), \qquad \ldots\ldots(6.31)$$

where Q_N is defined as $\lambda^{3N} Z_N$. This formula can be compared with the classical formula (5.3). The normalizing condition (6.7) can now be written as

$$Q_N = \frac{1}{N!} \int W_N(\mathbf{r}^N)\, d\mathbf{r}^N, \qquad \ldots\ldots(6.32)$$

which formulae is again in agreement with the classical expression (5.4 *a*), Q_N being the quantum mechanical configurational partition function, and by definition being related to the total partition function Z_N by $Q_N = Z_N \lambda^{3N}$.

The configurational probability density is defined in terms of $W_N(\mathbf{r}^N)$ by

$$n_h(\mathbf{r}^h) = \frac{N!}{(N-h)!} \int \mathscr{P}_N(\mathbf{r}^N;\ \mathbf{r}^N)\, d\mathbf{r}^{N-h} = \frac{1}{(N-h)!\ Q_N} \int W_N(\mathbf{r}^N)\, d\mathbf{r}^{N-h}, \qquad \ldots\ldots(6.33)$$

which expression is the exact analogue of the corresponding classical relation (5.5). So the Slater sum $W_N(\mathbf{r}^N)$ takes in quantum theory in many respects the important part played by the Boltzmann factor $\exp(-\Phi(\mathbf{r}^N)/kT)$ in classical theory.

§7. SERIES EXPANSION IN POWERS OF THE DENSITY OF DISTRIBUTION FUNCTION AND EQUATION OF STATE OF COMPRESSED GASES

(i) *Introduction*

In the previous sections, §§5 and 6, it has been shown that all important knowledge of the gaseous or liquid system can be derived from the configurational distribution function $\exp -\beta\Phi(\mathbf{r}^N)$, when classical theory can be applied, and from the quantum analogue, the Slater sum $W_N(\mathbf{r}^N)$, when quantum theory has to be applied. We shall base the considerations of this section on the general quantum mechanical Slater sum $W_N(\mathbf{r}^N)$ defined by (6.22) in order to make the results applicable also when quantum mechanics has to be applied.

In §§5 and 6 it has been shown also that the most important "reduced" distribution function (5.12), (6.33) which can be derived from $W_N(\mathbf{r}^N)$ is the pair distribution function

$$n_2(r_{12}) = (1/Q_N(N-2)!) \int \ldots \int W_N(\mathbf{r}^N)\, d\mathbf{r}_3 \ldots d\mathbf{r}_N.$$

This function has its own interest, being capable of direct experimental determination (cf. §10), and is of fundamental importance because of the fact that in classical theory as well as in quantum theory the thermal and caloric equations of state are expressible entirely in terms of this pair distribution function (cf. equations (5.8), (5.11), (6.17) and (6.18)).

On the other hand, integration of $W_N(\mathbf{r}^N)$ leads to the configurational partition function $Q_N = (1/N!) \int W_N(\mathbf{r}^N)\, d\mathbf{r}^N$ from which the total partition function $Z_N = Q_N/\lambda^{3N}$, the free energy $F = -kT \ln Z_N$ and thus the thermal and caloric equations of state can be obtained directly.

In this section these quantities: the reduced distribution function n_h, especially the pair distribution function n_2, the pressure pV and the energy U will be developed into power series of the number density n. For this purpose, these quantities are first expressed as a series in powers of the activity z (introduced in § 5 (iii)) which is approximately equal to the number density n at low densities. The equation needed to express all these quantities $n_h(z, T)$, $pV(z, T)$ and $U(z, T)$ in terms of n is an equation for z as function of n and T. This equation is obtained as a particular case of the equations $n_h(z, T)$ for $h=1$.

(ii) *Development of Pressure and Energy in Powers of z*

To obtain the development of pressure and energy into a power series of z, it is necessary to transform $W_N(\mathbf{r}^N)$, by introducing U-functions defined in the following way (cf. Beth and Uhlenbeck 1937, Kahn 1938):

$$\left.\begin{aligned} W_1(\mathbf{r}_\iota) &= U_1(\mathbf{r}_\iota) \\ W_2(\mathbf{r}_\iota\mathbf{r}_\kappa) &= U_2(\mathbf{r}_\iota\mathbf{r}_\kappa) + U_1(\mathbf{r}_\iota)U_1(\mathbf{r}_\kappa) \\ W_3(\mathbf{r}_\iota\mathbf{r}_\kappa\mathbf{r}_\lambda) &= U_3(\mathbf{r}_\iota\mathbf{r}_\kappa\mathbf{r}_\lambda) + \Sigma_{\mathrm{P}} U_2(\mathbf{r}_\iota\mathbf{r}_\kappa)U_1(\mathbf{r}_\lambda) + U_1(\mathbf{r}_\iota)U_1(\mathbf{r}_\kappa)U_1(\mathbf{r}_\lambda) \end{aligned}\right\}, \quad \ldots\ldots (7.1)$$

where $W_h(\mathbf{r}_1 \ldots \mathbf{r}_h)$ is the Slater sum, defined by (6.22), assuming the molecules 1 to h to be present alone in the volume V. The equations (7.1) can be solved in the form:

$$\left.\begin{aligned} U_1(\mathbf{r}_\kappa) &= W_1(\mathbf{r}_\iota) \\ U_2(\mathbf{r}_\iota\mathbf{r}_\kappa) &= W_2(\mathbf{r}_\iota\mathbf{r}_\kappa) - W_1(\mathbf{r}_\iota)W_1(\mathbf{r}_\kappa) \\ U_3(\mathbf{r}_\iota\mathbf{r}_\kappa\mathbf{r}_\lambda) &= W_3(\mathbf{r}_\iota\mathbf{r}_\kappa\mathbf{r}_\lambda) - \Sigma_{\mathrm{P}} W_2(\mathbf{r}_\iota\mathbf{r}_\kappa)W_1(\mathbf{r}_\lambda) + 2W_1(\mathbf{r}_\iota)W_1(\mathbf{r}_\kappa)W_1(\mathbf{r}_\lambda) \end{aligned}\right\}. \quad \ldots\ldots (7.2)$$

The expressions of W in terms of U and those of U in terms of W are exactly the same except for the coefficient $(-1)^{k-1}(k-1)!$, where k is the number of groups in the term. In general the expression of $W_l(\mathbf{r}_1 \ldots \mathbf{r}_l)$ is obtained by dividing the l molecules in all possible ways in groups, making the corresponding product of U functions and summing this over all possible divisions. So in general

$$W_N(\mathbf{r}^N) = \mathbf{S\Pi} U_l(\mathbf{r}^l), \quad \ldots\ldots (7.3)$$

where $\mathbf{S\Pi}$ denotes a sum of products of U functions over all possible divisions of the N molecules in separate groups. The U_l functions have the important property that *they approach zero for a "separate" configuration, in which the l molecules are divided into two or more groups which are separated from each other and beyond the range of intermolecular forces.* This property can for small l directly be verified from the definitions (7.2).

This transformation can be obtained in classical statistics in a little different way by generalizing Ursell's (1947) procedure, as has been shown by Mayer and collaborators (1937–1938). It has the disadvantage, however, that classical statistics have to be assumed and that the assumption of the additivity of intermolecular forces has to be introduced at an early stage:

(*a*) In classical statistics $W_N(\mathbf{r}^N) = \exp\{-\beta\Phi(\mathbf{r}^N)\}$.

(*b*) When the intermolecular forces are additive this transforms into $\exp -\beta \sum_{\iota>\kappa}\sum_{\kappa} \phi(r_{\iota\kappa})$.

The expression for $W_N(\mathbf{r}^N)$ can then be transformed into the identity

$$W_N(\mathbf{r}^N) = \prod_{\iota>\kappa} \exp\{-\beta\phi(r_{\iota\kappa})\} = \prod_{\iota>\kappa} (1+f_{\iota\kappa}), \quad \ldots\ldots (7.4)$$

where for classical statistics, $f_{\iota\kappa}=\exp\{-\beta\phi(r_{\iota\kappa})\}-1$ is identical with $U_2(\mathbf{r}_\iota\mathbf{r}_\kappa)$. Comparing this transformation with the previous one, the following expressions for the U_l functions are found:

$$U_1(\mathbf{r}_\iota)=1; \qquad U_2(\mathbf{r}_\iota\mathbf{r}_\iota)=f_{\iota\kappa}; \qquad U_3(\mathbf{r}_\iota\mathbf{r}_\kappa\mathbf{r}_\lambda)=f_{\iota\kappa}f_{\kappa\lambda}f_{\lambda\mu}+\Sigma f_{\iota\kappa}f_{\kappa\lambda}; \quad \ldots\ldots(7.5)$$

where the U_l function now contains all products of $f_{\iota\kappa}$ each of which must connect together at least singly all l molecules. The basic property of the U_l functions, that they approach zero for a "separated" configuration can easily be understood, since then always at least one of the factors $f_{\iota\kappa}$ connecting a molecule of one group with a molecule of the other group is zero. The condition that all l molecules must be in a configuration so that at least one product $f_{\iota\kappa}f_{\kappa\lambda}\ldots$. connecting all l molecules is different from zero, is expressed by the expression that the molecules form a "cluster".

Integration of the Slater sum (7.3) over all the coordinates leads to integrals of the type

$$b_l=\frac{1}{l!}\int\ldots\int U_l(\mathbf{r}_1\mathbf{r}_2\ldots.\mathbf{r}_l)\,d\mathbf{r}_2\,d\mathbf{r}_3\ldots.d\mathbf{r}_l. \qquad \ldots\ldots(7.6)$$

As the integrand is zero for a "separate" configuration, integration over the coordinates of the molecules $2,\ldots.,\ l$ relative to 1 leads to $l!\,b_l$, which result is independent of the coordinate $\mathbf{r}_1$. These integrals, called "*cluster integrals*", play an important part in the theory of the equation of state. In classical theory these cluster integrals can be evaluated in principle if the intermolecular field $\phi(r_{\iota\kappa})$ is known, although evaluation even of b_3 offers many difficulties. When quantum theory has to be applied, the evaluation of these cluster integrals offers still more difficulties, as then the difficult quantum mechanical problem is superposed on the mathematical difficulties which exist already in classical theory.

Integration of (7.3) over all coordinates gives the configurational partition function (5.4) from which, using $\lambda^{3N}Z_N=Q_N$, the total partition function Z_N can be obtained. The result is

$$Q_N=\frac{1}{N!}\int W_N(\mathbf{r}^N)\,d\mathbf{r}^N=\mathbf{S}\Pi_{(\Sigma lm_l=N)}\frac{(Vbl)^{m_l}}{m_l!}, \qquad \ldots\ldots(7.7)$$

where the summation $\mathbf{S}$ must be carried out over all possible divisions of the N molecules into m_1 groups of 1, m_2 groups of 2, ..., m_l groups of l, etc. molecules, so that $\Sigma lm_l=N$.

The evaluation of this configurational partition function can now be carried out by applying the method of steepest descent (Kahn and Uhlenbeck 1938, Born and Fuchs 1938), leading to $\ln Q_N=\Sigma Vb_l z^l-N\ln z$. Application of equations (5.16) and (5.17) then gives the caloric and thermal equations of state. We follow here another method (de Boer 1940), which is simpler, but is limited to the region where the resulting power series converges.

The thermal equation of state can be obtained directly by applying the formula $p=kT\partial\ln Z_N/\partial V=kT\partial\ln Q_N/\partial V$ (cf. (5.17)) to expression (7.5). Differentiation with respect to V gives

$$pV=kT\Sigma_l Vb_l\frac{Q_{N-l}}{Q_N}=kT\Sigma_l Vb_l z^l, \qquad \ldots\ldots(7.8)$$

if the definition (5.20) of the activity and the fact that $l\ll N$ are taken into account.

The caloric equation of state is obtained from the formula (5.16),

$$U=kT^2\partial\ln Z_N/\partial T=\tfrac{3}{2}NkT+kT^2\partial\ln Q_N/\partial T,$$

which gives

$$U=\tfrac{3}{2}NkT+kT^2\Sigma_l\frac{dVb_l}{dT}\frac{Q_{N-1}}{Q_N}=\tfrac{3}{2}NkT+kT^2\Sigma_l\frac{dVb_l}{dT}z^l. \quad \ldots\ldots(7.9)$$

Thus both the equations (7.6) and (7.7) are obtained in the form of a series expansion into powers of z. The coefficients b_l can be calculated in principle as functions of temperature from their definition (7.3). The proper expansions of pV and U in terms of the number density n can be found when the activity is known as function of T and V. This equation is obtained in the next section.

(iii) *Development of Distribution Functions $n_h(\mathbf{r}^h)$ in powers of z.*

The transformation of the Slater sum introduced in the previous section has been made to arrange the contributions to the partition function in such a manner that a series expansion in powers of the activity is obtained.

The integral expressions for the reduced distribution functions differ from the partition function $Q_N=(1/N!)\int W_N(\mathbf{r}^N)\,d\mathbf{r}_1\ldots d\mathbf{r}_N$ in so far as the integration for the evaluation of $n_h(\mathbf{r}_1,\ldots,\mathbf{r}_h)$ is carried out only over a limited number of molecules: over all molecules except the h molecules whose distribution is investigated $n_h=(1/Q_N(N-h)!)\int W_N(\mathbf{r}^N)\,d\mathbf{r}_{h+1}\ldots d\mathbf{r}_N$. Therefore it is suitable to start with a transformation of $W_N(\mathbf{r}^N)$ which differs slightly from the one used in the previous section. The transformation of the Slater sum $W_N(\mathbf{r}^N)$ into the U functions can be maintained only if the group $\{h\}$ is *considered formally as one molecule, with coordinate* $\mathbf{r}^h$. The definition of all U functions not containing $\{h\}$, according to (7.2), remains exactly the same. The U functions containing $\{h\}$ are defined by similar relations:

$$\left.\begin{aligned} W(\mathbf{r}^h)&=U(\mathbf{r}^h),\\ W(\mathbf{r}^h\mathbf{r}_\kappa)&=U(\mathbf{r}^h\mathbf{r}_\kappa)+U(\mathbf{r}^h)U_1(\mathbf{r}_\kappa),\\ W(\mathbf{r}^h\mathbf{r}_\kappa\mathbf{r}_\lambda)&=U(\mathbf{r}^h\mathbf{r}_\kappa\mathbf{r}_\lambda)+\Sigma U(\mathbf{r}^h\mathbf{r}_\kappa)U_1(\mathbf{r}_\lambda)+U(\mathbf{r}^h)U_1(\mathbf{r}_\kappa)U_1(\mathbf{r}_\lambda),\ \text{etc.}\end{aligned}\right\} \quad \ldots\ldots(7.10)$$

which can be solved in the form

$$\left.\begin{aligned} U(\mathbf{r}^h)&=W(\mathbf{r}^h),\\ U(\mathbf{r}^h\mathbf{r}_\kappa)&=W(\mathbf{r}^h\mathbf{r}_\kappa)-W(\mathbf{r}^h)W_1(\mathbf{r}_\kappa),\\ U(\mathbf{r}^h\mathbf{r}_\kappa\mathbf{r}_\lambda)&=W(\mathbf{r}^h\mathbf{r}_\kappa\mathbf{r}_\lambda)-\Sigma W(\mathbf{r}^h\mathbf{r}_\kappa)W_1(\mathbf{r}_\lambda)+2W(\mathbf{r}^h)W_1(\mathbf{r}_\kappa)W_1(\mathbf{r}_\lambda),\ \text{etc.}\end{aligned}\right\} \quad \ldots\ldots(7.11)$$

The above equations are exactly the same as those of (7.1) and (7.2), except that everywhere $\mathbf{r}^h$ is substituted in place of $\mathbf{r}_i$. In general $W(\mathbf{r}^h\mathbf{r}_{h+1}\ldots\mathbf{r}_N)$ is given as a sum over all products of U functions, similar to (7.3). The only difference is that in this sum there occur U functions of the type (7.2) and in each term always one of the type (7.11). It is possible to arrange these terms in the following way:

$$\begin{aligned} W_N(\mathbf{r}^N)=U_1(\mathbf{r}^h)W_{N-h}(\mathbf{r}_{h+1}\ldots\mathbf{r}_N)+\Sigma_\kappa U_2(\mathbf{r}^h\mathbf{r}_\kappa)W_{N-h-1}\\ +\Sigma_\kappa\Sigma_\lambda U_3(\mathbf{r}^h\mathbf{r}_\kappa\mathbf{r}_\lambda)W_{N-h-2}+\ldots. \end{aligned} \quad \ldots\ldots(7.12)$$

From this series expansion the series expansion for $n_h(\mathbf{r}^h)$ can be obtained quite easily, by integration of (7.12) over $\mathbf{r}_{h+1},\ldots,\mathbf{r}_N$. Then new integrals appear analogous to the cluster integrals (7.6), viz.:

$$b_l^{(h)}(\mathbf{r}^h)=\frac{1}{l!}\int U(\mathbf{r}^h\mathbf{r}_{h+1}\ldots\mathbf{r}_{h+l-1})\,d\mathbf{r}_{h+1}\ldots d\mathbf{r}_{h+l-1}. \quad \ldots\ldots(7.13)$$

These cluster integrals depend on the relative coordinates $\mathbf{r}^h$ of the group $\{h\}$. For $h=1$ the $b_l^{(1)}$ are identical again with the cluster integrals (7.6): $b_l^{(1)}\equiv b_l$. Using these integrals $b_l^{(h)}$, the following expression is obtained directly by integrating (7.12) over $\mathbf{r}_{h+1}\ldots\mathbf{r}_N$:

$$n_h(\mathbf{r}^h)=\frac{1}{Q_N(N-h)!}\int W_N(\mathbf{r}^N)\,d\mathbf{r}_{h+1}\ldots d\mathbf{r}_N=\sum_{l=1}^{\infty} l b_l^{(h)}\frac{Q_{N-h-l+1}}{Q_N}. \quad \ldots\ldots(7.14)$$

Introducing the activity $z=Q_{N-1}/Q_N$ as in § 5 (iii) then gives

$$n_h(\mathbf{r}^h)=z^{h-1}\sum_1^{\infty} l b_l^{(h)}(\mathbf{r}^h)z^l. \quad \ldots\ldots(7.15)$$

This is the general expression for the development of $n_h(\mathbf{r}^h)$ in powers of z; the coefficients are the integrals $b_l^{(h)}(\mathbf{r}^h)$, which depend on the configuration of the h molecules and which are defined by (7.13) and (7.11). Here we shall not longer consider this general form but will apply it directly to the most important cases $h=2$ and $h=1$. For $h=2$ one obtains

$$n_2(r_{12})=\sum_1^{\infty} l b_l^{(2)}(r_{12})z^{l+1}, \quad \ldots\ldots(7.16)$$

the first coefficients being

$$\left.\begin{aligned} b_1^{(2)}(r_{12}) &= W(\mathbf{r}_1,\mathbf{r}_2),\\ 2!\,b_2^{(2)}(r_{12}) &= \int[W(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)-W(\mathbf{r}_1,\mathbf{r}_2)W(\mathbf{r}_3)]\,d\mathbf{r}_3, \text{ etc.}\end{aligned}\right\} \quad \ldots\ldots(7.17)$$

Since $b_l^{(1)}\equiv b_l$, one obtains, for $h=1$,

$$n_1=n=\sum_1^{\infty} l b_l z^l. \quad \ldots\ldots(7.18)$$

For simplicity it is assumed that no external field of force is present; this makes n_2 only dependent on r_{12} and n_1 *in*dependent of $\mathbf{r}_1$ and equal to n.

This second equation for n in terms of z gives us at once the possibility of obtaining the expression for z in terms of n. Born and Fuchs (1938) have shown that the equation can be solved in the form of an exponential function

$$z=n\exp(-\Sigma_k\beta_k n^k), \quad \ldots\ldots(7.19)$$

where

$$\beta_1=2b_2;\quad \beta_2=3b_3-6b_2^2;\quad \beta_3=4b_4-24b_2b_3+\frac{80}{3}b_2^3;\ \text{etc.} \quad \ldots\ldots(7.20)$$

These expressions can also be formed by expanding z into a power series of n, substituting in (7.18), and equating to zero terms with equal powers of n. Thus for low densities *the activity approaches simply the number density n* of molecules. At higher densities, where the terms in the exponent cannot be neglected, large deviations may occur.

In classical theory the integrals β_k have a simple meaning being the so-called " irreducible integrals " introduced by Mayer. The simplest are :

$$\left.\begin{aligned} V\beta_1 &= \iint f_{12}\,d\mathbf{r}_1\,d\mathbf{r}_2,\\ 2!\,V\beta_2 &= \iiint f_{13}f_{23}f_{31}\,d\mathbf{r}_1\,d\mathbf{r}_2\,d\mathbf{r}_3,\\ 3!\,V\beta_3 &= \iiiint (f_{13}f_{13}f_{14}f_{23}f_{34}f_{34}+6f_{12}f_{23}f_{34}f_{12}f_{14}+3f_{12}f_{23}f_{34}f_{41})\,d\mathbf{r}_1\,d\mathbf{r}_2\,d\mathbf{r}_3\,d\mathbf{r}_4.\end{aligned}\right\} \quad \ldots\ldots(7.21)$$

The ordinary cluster integrals are " reducible " in the sense that they can often be written as a product of two or more integrals. Thus for instance $\int f_{12}f_{13}\,d\mathbf{r}_1\,d\mathbf{r}_2\,d\mathbf{r}_3=V\int f_{12}\,d\mathbf{r}_2\int f_{13}\,d\mathbf{r}_3$. The irreducible integrals in which the molecules of the cluster are connected by more than one series of $f_{\alpha\kappa}$ cannot be factorized in this way. For more details the reader is referred to Born and Fuchs (1938) and to the monograph *Statistical Mechanics* (1940) of Mayer and Mayer. In quantum theory no such interpretation of the β_k seems to be possible.

The method used in this section to derive the general expressions for the distribution function is a generalization of Uhlenbeck's U function method, which has not been published so far. A slightly more elaborate method, which is also based on the U function method has been used by de Boer (1940). The advantage of the use of these U functions is that the results are applicable in quantum theory as well as in classical theory.

If classical theory is assumed to be valid from the beginning and the intermolecular forces are assumed to be additive, the results of this section can also be obtained, as has been mentioned with the f_{ik} method of Ursell. This method seems to have been applied to the calculation of the distribution function first by Yvon in 1935, in an attempt to derive quite generally the equation of state for a gas subjected to the external gravitational field. In 1937 Yvon paid general attention to the connection between the fluctuations in density and the distribution function, applying again the Ursell development for the distribution functions (cf. § 9). The results which have been obtained independently, using the general U function method of Uhlenbeck, by de Boer (1939, 1940) agreed with these results of Yvon. In 1941 Mayer and Montroll, basing their work upon the classical f_{ik} method of Ursell, derived general expressions for the distribution functions $n_k(\mathbf{r}^k)$, which can be derived from the expressions (7.15) and (7.16) given in this section by introducing the assumptions of classical theory and of additivity of molecular forces.

(iv) *Power Series for pV, U and $n_2(r_{12})$ in powers of n.*

The substitution of the series expansions (7.19) obtained in sub-section (iii) for z in the expressions (7.9), (7.10) and (7.16) gives the following power series for pV, U and $n_2(r_{12})$:

$$pV = RT - RT\Sigma \frac{k}{k+1} \beta_k n^k, \qquad \ldots\ldots (7.22)$$

$$U = \tfrac{3}{2} RT + RT^2 \Sigma \frac{1}{k+1} \frac{d\beta_k}{dT} n^k, \qquad \ldots\ldots (7.23)$$

$$n_2(r_{12}) = \Sigma_k \gamma_k(r_{12}) n^k. \qquad \ldots\ldots (7.24)$$

The coefficients $\gamma_k(r_{12})$ can easily be obtained, the first ones being given by

$$\left.\begin{aligned} \gamma_2(r_{12}) &= b_1^{(2)}(r_{12}), \\ \gamma_3(r_{12}) &= 2b_2^{(2)}(r_{12}) - 4b_2 b_1^{(2)}(r_{12}), \\ \gamma_4(r_{12}) &= 3b_3^{(2)}(r_{12}) - 12b_2 b_2^{(2)}(r_{12}) + (4b_2^2 - 6b_3) b_1^{(2)}(r_{12}), \text{ etc.} \end{aligned}\right\} \qquad \ldots\ldots (7.25)$$

Comparison of these formulae with the experimental data is postponed to § 8 (equation of state) and to § 9 (distribution functions).

These expressions for the caloric and thermal equations of state only converge when the number density is small. Experimentally it has been possible to express these quantities as a power series of the density only at relatively low densities (cf. § 8 (ii)) which may be said to be about 100 to 200 times the gas-density at 1 atmosphere, provided the temperature is not too close to the critical temperature. A discussion of the theoretical problems connected with the condensation and the critical phenomena lies beyond the scope of this survey. For this the interested reader is referred to the considerations of Born and Fuchs (1938), Kahn and Uhlenbeck (1938), Kahn (1938) and to the monograph of Mayer and Mayer (1940).

§8. CALCULATION OF THE EQUATION OF STATE

(i) *Law of Corresponding States*

For the further evaluation of the equation of state given by expression (7.22) of the previous section, it is necessary to make more definite assumptions as to the precise form of the intermolecular field which is the potential of the intermolecular forces. Before going into details of this field, it is of advantage to consider it from a somewhat more general point of view, in order to show that under certain conditions it can be proved that a law of corresponding states holds, which makes it possible to give the equation of state of different substances the same form, by reducing the macroscopic variables pressure, volume and temperature in a suitable way.

First consider the case of *classical theory*. Then according to the general principles of statistical mechanics given in § 5 (iii), the evaluation of the pressure is reduced to the evaluation of the configurational partition function Q_N which in classical theory takes the form $(1/N!)\exp\int(-\Phi/kT)\,d\mathbf{r}^N$ and from which the pressure follows by the formula $p=-kT\partial\ln Q_N/\partial V$. Now a law of corresponding states can be derived provided the following conditions are satisfied:

(*a*) The potential field $\Phi(\mathbf{r}^N)$ consists additively of the contributions of the different pairs of molecules: $\Phi(\mathbf{r}^N)=\frac{1}{2}\Sigma_\iota\Sigma_\kappa\phi(r_{\iota\kappa})$ (*assumption of additivity*).

(*b*) The intermolecular field $\phi(r_{\iota\kappa})$ between two molecules has for all substances the same form $\epsilon f(r_{\iota\kappa}/\sigma)=\phi(r_{\iota\kappa})$, except for the two scale factors ϵ and σ; ϵ has the dimension of an energy and σ the dimension of a length. So the function $f(x)$ should be the same for all substances, the only difference being the values of ϵ and σ (*assumption of similarity*).

If these two conditions are satisfied it is easily seen that, introducing $\mathbf{r}^*=\mathbf{r}/\sigma$, Q_N can be written in the form

$$Q_N=\frac{\sigma^{3N}}{N!}\int\ldots\int\exp-\sum\sum_{\iota>\kappa}\frac{\epsilon f(r^*_{\iota\kappa})}{kT}\,d\mathbf{r}^*_1\ldots d\mathbf{r}^*_N=\frac{\sigma^{3N}}{N!}J(T^*,V^*)^N. \qquad \ldots\ldots(8.1)$$

This integral expression depends only on the value of the reduced temperature $T^*=kT/\epsilon$ and of the reduced volume $V^*=V/N\sigma^3$. If now p is expressed in ϵ/σ^3 as a unit the following equation holds:

$$p^*=-\frac{1}{N}T^*\frac{\partial\ln Q_N}{\partial V^*}=-T^*\frac{\partial\ln J(T^*,V^*)}{\partial V^*}, \qquad \ldots\ldots(8.2)$$

which proves that, using the units ϵ/κ, $N\sigma^3$ and ϵ/σ^3 for temperature, molar volume and pressure, the equation of state takes the form of a law of corresponding states. These units are units derived from the scale factors occurring in the law of interaction of the molecules. The same can easily be verified from the expression (7.22): each β_k has the dimension $[\text{volume}]^k$, so that $\beta_k^*(T^*)=\beta_k\sigma^{-3k}$. If we introduce reduced densities $n^*=n\sigma^3$ as unit, then n^* is the average number of molecules in σ^3 and is the reciprocal of $V^*=V/n\sigma^3=n^{*-1}$, and the equation of state is given by

$$p^*V^*=T^*-T^*\sum\frac{k}{k+1}\beta_k^*n^{*k}. \qquad \ldots\ldots(8.3)$$

Of course this form of the law of corresponding states leads to the original form of Van der Waals: The critical data V_{cr}, p_{cr}, and T_{cr} expressed in these molecular units: i.e. $V^*_{\text{cr}}=V_{\text{cr}}/N\sigma^3$, $p^*_{\text{cr}}=p_{\text{cr}}\sigma^3/\epsilon$ and $T^*_{\text{cr}}=kT_{\text{cr}}/\epsilon$ should have the

same value for all substances. Consequently we not only have $p^* = p^*(V^*T^*)$ but also $p^*/p^*_{cr} = f(V^*/V^*_{cr}, T^*/T^*_{cr})$ or $p/p_{cr} = f(V/V_{cr}, T/T_{cr})$ and the same would happen in reducing the pressure, temperature and molar volume by any other characteristic point of the equation of state.

In the case of *quantum statistical mechanics* the situation is quite different. Then again the equation of state can be deduced from Q_N, but now Q_N can be written as

$$Q_N = (\sigma\lambda)^{3N} \int \ldots \int \left(\Sigma_\varrho \phi_\varrho(\mathbf{r}^{N*}) \exp\left(-\frac{\mathscr{H}^*}{T^*} \right) \phi_\varrho(\mathbf{r}^{N*}) \right) d\mathbf{r}^{N*}. \quad \ldots\ldots(8.4)$$

The functions ϕ_ϱ depend on $r_1^*, \ldots\ldots r_N^*$ and on the reduced volume V^*, and the reduced Hamiltonian operator $\mathscr{H}^* = \mathscr{H}/\epsilon$ takes the form

$$\mathscr{H}^*(r^{N*}, \Lambda^*) = -\Lambda^{*2} \frac{1}{8\pi^2} \Sigma_i \frac{\partial}{\partial \mathbf{r}_i^*} \cdot \frac{\partial}{\partial \mathbf{r}_i^*} + \sum_{i>k} \sum f(r_{ik}^*), \quad \ldots\ldots(8.5)$$

where Λ^{*2} is written for the factor $h^2/\sigma^2 m\epsilon$. So the Hamiltonian operator does not only depend on $\mathbf{r}^* = \mathbf{r}/\sigma$ but contains also Λ^* as a factor. This fact destroys the validity of the law of corresponding states, because now $Q_N = \sigma^{3N} J(T^*, V^*, \Lambda^*)^N$ and consequently p^* becomes

$$p^* = -T^* \frac{\partial \ln J(T^*, V^*, \Lambda^*)}{\partial V^*}. \quad \ldots\ldots(8.6)$$

Therefore the reduced pressure depends not only on the reduced temperature and the reduced volume but also on the quantity Λ^* as a parameter. This parameter Λ^* varies from substance to substance since it is a function of the scale factors ϵ and σ and of the *mass* of the molecules. Its physical meaning is clear: It is necessary that a new length Λ be introduced into the theory which is closely connected with the wave character of the molecules. This quantity $\Lambda = h/(m\epsilon)^{\frac{1}{2}}$ is the de Broglie wavelength of the relative motion of two molecules with relative kinetic energy ϵ. Evidently the ratio of this characteristic length Λ to the length σ characterizing the dimensions of the molecular field should occur in the quantum theory. This ratio is $\Lambda^* = \Lambda/\sigma = h/\sigma(m\epsilon)^{\frac{1}{2}}$, the quantum mechanical parameter occurring in the equation of state (8.6) in quantum theory.

In general the effects of quantum theory on the equation of state will be small, when Λ^* is small, as then the de Broglie wavelength is small compared with the scale factor characterizing the intermolecular potential field and is thus small compared to the molecules. As in optics, when in cases where λ is small compared with the dimensions of the objects geometrical optics can be used, so here classical statistical mechanics can be used when Λ^* is small. This will appear to be the case for gases like A, N_2, Ne, Kr and Xe. For the lighter gases, especially deuterium, hydrogen and helium, Λ^* proves to be larger than unity, with the consequence that in these cases quantum effects can no longer be neglected and the quantum mechanical expressions of the equation of state are to be used.

This determination of the law of corresponding states in terms of the "molecular units" given by de Boer and Michels (1938) has been stated in classical theory by Pitzer (1939) and Guggenheim (1945) and has been given in this form by de Boer (1948), showing the possibility of using it in the quantum mechanical case as a tool to predict properties of unknown substances and to study the influence of quantum theory on the properties of matter. As an example of the first may be mentioned the prediction of the properties of the light isotope of helium

(de Boer and Lunbeck 1948 b) and of the second, the studies of the properties of the solid (de Boer and Blaisse 1948) and of the liquid state (de Boer and Lunbeck 1949 a) of the permanent and noble gases. Finally it may be mentioned that as early as 1921 Byk pointed out the necessity of accepting deviations from the law of corresponding states as a result of the part played by the elementary quantum of action h in the behaviour of light substances. These considerations were, however, based on the van de Waals' theory of the law of corresponding states.

(ii) *The Experimental Data*

The experimental knowledge of the properties of compressed gases is chiefly limited to the numerous results of the determinations of compressibilities of gases and to the measurements of the Joule-Thomson coefficients. The measurements of the *isotherms of gases* are all expressed in the form of a power series:

$$pV = RT\left(1 + \frac{B(T)}{V} + \frac{C(T)}{V} + \ldots.\right), \qquad \ldots\ldots(8.7)$$

where the second, third, etc. virial coefficients $B(T)$, $C(T)$, etc. can be chosen so that expression (8.7) represents the experimental data at each temperature. At higher density it seems to be impossible to represent the data with sufficient accuracy by a series of this type (cf. Michels and Michels (1937 a, b), CO_2; Michels, Wouters and de Boer (1936 a), N_2; Michels and Goudeket (1941 a, b), H_2, D_2; Michels and Geldermans (1942)).

The measurements of the *Joule-Thomson coefficient*, defined thermodynamically by the change of temperature with pressure at constant enthalpy $\mu = (\partial T/\partial p)_H$, have been carried out mainly by Roebuck and Osterberg (1933–1935). The results are also expressed in the form of a power series:

$$\mu = \mu_0(T) + \mu_1(T)p + \ldots. \qquad \ldots\ldots(8.8)$$

These two expressions are related to each other by thermodynamic relations. The Joule-Thomson coefficient can be written as $\mu = -(1/c_p)(\partial H/\partial p)_T$. On the other hand $(\partial H/\partial p)_T = -T(\partial V/\partial T)_p + V$, which can be calculated from (8.7) giving

$$\mu_0(T) = \frac{1}{c_{p0}}\left(T\frac{\partial B}{\partial T} - B\right) \qquad \ldots\ldots(8.9)$$

and similar relations for $\mu_1(T)$, where c_{p0} is the first term of $c_p = c_{p0} + c_{p1}p + \ldots.$ This shows the possibility of investigating the experimental consistency of the compressibility measurements with those of the Joule-Thomson coefficient (cf. de Groot and Geldermans (1947), de Groot and Michels (1948)).

Experimental determinations of the *caloric equation of state* of compressed gases seem to be very scarce. Only for some gases and vapours of interest for cooling technique and steam engines have measurements been carried out. On the other hand, Michels and collaborators (Michels, Wouters and de Boer (1936 b), Michels, Bijl and Michels (1937 c), Michels and Goudeket (1941 c), have calculated the caloric equation of state by using the thermodynamic relation $(\partial U/\partial V)_T = T(\partial p/\partial T)_p - p$, which, using (8.7) and integrating between $V = \infty$ and V, gives the caloric equation of state

$$U = U_0 - T\frac{\partial B}{\partial T}\frac{1}{V} - T\frac{\partial C}{\partial T}\frac{1}{V^2} - \ldots. \qquad \ldots\ldots(8.10)$$

The energy U_0 is the energy of the dilute gas, which can be measured or calculated directly from spectroscopic data.

Therefore, when the measurements of the Joule-Thomson coefficients and of the second virial coefficient are thermodynamically consistent, only one of these coefficients needs really to be compared with the theoretical expressions.

As, however, the Joule-Thomson coefficient μ_0 is determined experimentally in an independent manner, the comparison of these results with theory gives an independent indication as to the form of the potential field. First we shall consider the experimental data of those substances which can be treated by classical theory. In section (iv) the substances which must be treated by quantum theory are considered more closely.

(iii) *Classical Theory*

A survey of the comparison between the theoretical and experimental data for the second virial coefficient has been given several times (cf. Lennard-Jones 1924, 1931, Fowler 1936, Chap. x., Lennard-Jones 1937) and the reader is referred to these original articles for all details. The *second virial coefficient* $B(T)$ is, (7.22),

$$B = -\frac{N}{2}\beta_1 = -Nb_2 = -\frac{N}{2}\int (\exp\{-\beta\phi(r_{12})\} - 1)\, d\mathbf{r}_2. \quad \ldots\ldots(8.11)$$

The theoretical expression to be substituted for $\phi(r)$ is taken to be of the form

$$\phi(r) = \epsilon\kappa(s)\left[\left(\frac{\sigma}{r}\right)^s - \left(\frac{\sigma}{r}\right)^6\right]. \quad \ldots\ldots(8.12)$$

The attractive part of the interaction energy $\phi(r)$ is taken to be proportional to r^{-6}, in accordance with the Heitler–London theory of intermolecular forces. The repulsive potential field is taken to be proportional to r^{-s}, where the exponent s lies between 9 and 14. In Figure 1, this potential field is given

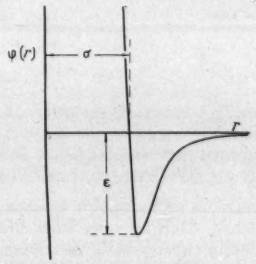


Figure 1. Molecular potential field $\phi(r) = 4\epsilon\{(r/\sigma)^{-12} - (r/\sigma)^{-6}\}$

for $s = 12$. Clearly σ has the physical meaning of distance for which $\phi(r) = 0$. If the quantity ϵ is taken to be the energy in the potential minimum, $\kappa(s)$ is only a numerical factor, having the value $(6/s-6)(s/6)^{s/s-6}$. This expression for the intermolecular field must now be substituted in (8.11). The integration is carried out and $B(T)$ can be calculated as a function of temperature.

From a comparison between theoretical values of $B(T)$ and the corresponding experimental values the values of the repulsive power exponent s and of ϵ and σ can be determined for each substance. For the substances neon, nitrogen and argon, relatively heavy gases which can be expected to show only small deviations from the classical behaviour, the value of s giving a $B(T)$ curve of a shape in agreement with the experimental data varies between 9 and 14; the value $s=12$ has therefore been chosen as a suitable one to be used in this article. The constant $\kappa(s)$ then takes the value $\kappa=4$. Substitution of (8.12) in (8.11) then leads to

$$B^*(T^*)=\frac{2\pi}{3}\left(\frac{4}{T^*}\right)^{\frac{1}{4}}\sum_{n=0}^{\infty}c_n\left(\frac{4}{T^*}\right)^{n/2}, \qquad \ldots\ldots(8.13)$$

where $B^*=B/N\sigma^3$, $c_n=-\Gamma(\frac{1}{2}n-\frac{1}{4})/4n!$ and $T^*=kT/\epsilon$. At high temperatures the sum becomes constant and B^* becomes proportional to $T^{*-1/4}$. For a gas of

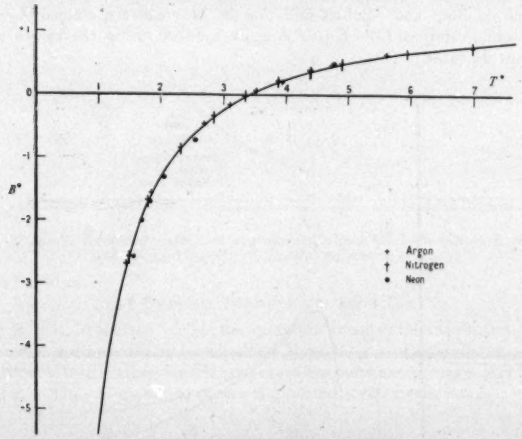


Figure 2. Classical theoretical curve and experimental points for second virial coefficient $B^*=B/(N\sigma^3)$ as function of $T^*=kT/\epsilon$.

elastic spheres B^* should be equal to $2\pi/3$. The slight dependence on T^* results from the fact that hard spheres are not used, but a repulsive potential field proportional to r^{-12}. If now log B^* is plotted as a function of log T^* on transparent paper and the experimental values of log B are also plotted as a function of log T on another sheet of paper, then, if the two plots coincide, the difference in abscissa gives log $N\sigma^3$ and the difference in ordinate gives log ϵ/κ. The values of ϵ and σ so obtained by Lennard-Jones (1924) are given in the table. In Figure 2 the

Gas	ϵ/k (° K.)	$N\sigma^3$ (cm^3)	ϵ (10^{-16} ergs)	σ (A.)
Neon	35·6	12·50	49·3	2·74
Argon	120·3	24·0	167	3·42
Nitrogen	96·6	31·3	134	3·73

theoretical values of B^* have been plotted as a function of T^* together with the experimental values of A, N_2 and Ne, using the values of $N\sigma^3$ and ϵ/k given in the table in order to show that they fit the theoretical curve very well. The values of ϵ and σ of the last two columns are calculated from the values in the first, by taking the present values $N = 6{\cdot}02 \times 10^{23}$, $k = 1{\cdot}38 \times 10^{-16}$ from Birge (1942).

The values of the third *virial coefficient*

$$C = -\frac{2N^2}{3}\beta_2 = -\frac{N^2}{3}\int (\exp\{-\beta\phi(r_{12})\} - 1) \times (\exp\{-\beta\phi(r_{13})\} - 1)(\exp\{-\beta\phi(r_{23})\} - 1)\, d\mathbf{r}_2\, d\mathbf{r}_3 \quad \ldots\ldots (8.14)$$

have been calculated, using the same potential field (8.12) with $s = 12$, $\kappa(s) = 4$ by de Boer and Michels (1939) by numerical integration, and by Montroll and Mayer (1941) by an approximation method based on Laplace transformation of the integral. The results of these calculations are reproduced in Figure 3 where the theoretical curve of de Boer and Michels and that of Montroll are compared with the experimental values of $C^* = C/(N\sigma^3)^2$, again reduced, using the values of σ and of ϵ from the table.

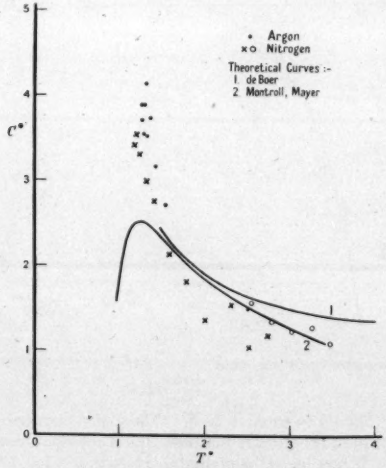


Figure 3. Classical theoretical curves and experimental points for third virial coeffiicent $C^* = C/(N\sigma^3)^2$ as function of $T^* = kT/\epsilon$.

The values of the *Joule–Thomson coefficient* $\mu_0(T)$ have been calculated theoretically by Hirschfelder, Ewell and Roebuck (1938), and Hirschfelder and Rosevaere (1939), using the potential field (8.12) introduced by Lennard-Jones. We again give the result only for $s = 12$; this is

$$C^*_{p0}\mu_0^* = \frac{2\pi}{3}\left(\frac{4}{T^*}\right)^{1/4} \sum_{n=0}^{\infty} d_n \left(\frac{4}{T^*}\right)^{n/2}, \quad \ldots\ldots (8.15)$$

where $d_n = \Gamma(\frac{1}{2}n - \frac{1}{4}).(5+2n)/16n!$, where $C^*_{p0}\mu^*_0 = C_{p0}\mu_0/N\sigma^3 = (C_{p0}/Nk)(\mu_0 k/\sigma^3)$ and where again $T^* = kT/\epsilon$. At high temperatures again $C^*_{p0}\mu^*_0$ approaches $T^{*-1/4}$, the sum approaching a constant value. In Figure 4 the theoretical values for N_2 and A are compared with the available experimental data.

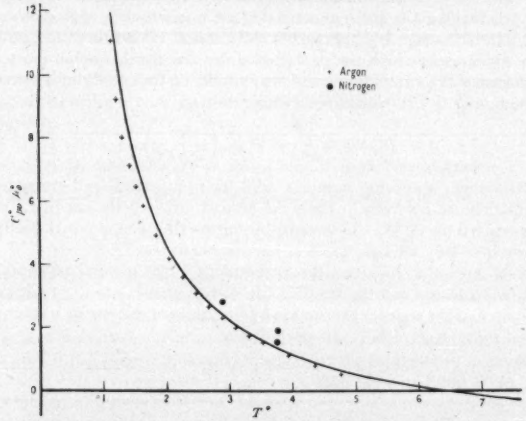


Figure 4. Classical theoretical curve and experimental points for Joule–Thomson coefficient $C^*_{p0}\mu^*_0 = (C_{p0}/R)(\mu_0 k/\sigma^3) = C_{p0}\mu_0/N\sigma^3$ as function of T^*.

(iv) *Quantum Theory of the Ideal Gas*

As has been discussed in §6 the quantum mechanical expression for the second virial coefficient can be obtained simply by replacing the expression $\exp\{-\beta\Phi(\mathbf{r}^N)\}$ by the corresponding quantum mechanical expression $W(\mathbf{r}^N)$ given by (6.22*b*). Thus in quantum theory the second virial coefficient is

$$B(T) = -Nb_2 = -\frac{N}{2}\int [W(r_{12}) - 1]d\mathbf{r}_2, \qquad \ldots\ldots(8.16)$$

where $W(r_{12}) = \lambda^6 2! \Sigma \phi^*_\varrho \exp(-\beta \mathscr{H}_{12})\phi_\varrho$, where $\mathscr{H}_{12}$ is the Hamiltonian operator of a pair of molecules.

The first important result of the replacement of $\exp(-\beta\Phi)$ by W is the fact that $B(T)$ is not zero even for an *ideal gas*. In fact the value of $W(r_1 r_2)$ for an ideal gas has been calculated in §6(iv) the results for two molecules being

$$W_{id}(r_{12}) = 1 \pm \exp\left(-\frac{2\pi}{\lambda^2} r^2_{12}\right). \qquad \ldots\ldots(8.17)$$

For molecules having symmetrical eigenfunctions the positive sign should be taken, for those with antisymmetrical eigenfunctions, the negative sign. According to the general argument of Ehrenfest and Oppenheimer (1931) molecules containing an even number of elementary particles (protons, neutrons and electrons) should have symmetrical eigenfunctions, whereas for those containing

an odd number of elementary particles the eigenfunctions should be antisymmetrical in the coordinates of the molecules. Most of the known molecules belong to the symmetrical case, the only exceptions of practical interest are the molecules HD and ^{3}He, both of which have antisymmetrical eigenfunctions.

In the case of symmetrical eigenfunctions W_{id} approaches the value 2 for $r_{12}=0$, whereas for the antisymmetrical eigenfunctions W_{id} approaches zero for $r_{12}=0$. In both cases W_{id} approaches the classical value unity for $r_{12}\gg\lambda$. The relative distances at which the deviations from the classical value unity become of importance are of the order of magnitude of the de Broglie wavelength. Introduction of (8.17) in (8.16) and integration gives

$$B_{id}=\mp N\frac{\lambda^3}{2^{5/2}}=\pm\frac{N}{2^{5/2}}\left(\frac{h^2}{2\pi mkT}\right)^{3/2}, \qquad \ldots\ldots(8.18)$$

where the upper and lower sign are valid for symmetrical and antisymmetrical eigenfunctions respectively. Thus the second virial coefficient of an ideal gas is proportional to $T^{-3/2}$. This contribution to the second virial coefficient is of importance only for light gases at low temperatures.

It is of interest to note that the introduction of the general expression (6.25) for $W_{id}(\mathbf{r}^N)$ into the expressions (7.2) for U and finally into (7.6) allows all the coefficients b_l of the series expansions in § 7 for pressure and energy to be integrated in closed form (Kahn 1938), the result being

$$b_l=\frac{(\pm1)^{l-1}}{l^{5/2}}\lambda^{3(l-1)}=\frac{(\pm1)^{l-1}}{l^{5/2}}\left(\frac{h^2}{2\pi mkT}\right)^{3(l-1)/2}. \qquad \ldots\ldots(8.19)$$

Introduction of these expressions in (7.8), (7.9) and (7.18) then gives the following general equations for the thermal and caloric equations of state and for the activity of the ideal gas:

$$p=\frac{kT}{\lambda^3}\sum_l\frac{(\pm1)^{l-1}}{l^{5/2}}(\lambda^3z)^l=kT\left(\frac{2\pi mkT}{h^2}\right)^{3/2}\sum_l\frac{(\pm1)^{l-1}}{l^{5/2}}\exp(l\mu/kT), \qquad \ldots\ldots(8.20)$$

$$U=\frac{3kT}{2\lambda^3}\sum_l\frac{(\pm1)^{l-1}}{l^{5/2}}(\lambda^3z)^l=\frac{3}{2}kT\left(\frac{2\pi mkT}{h^2}\right)^{3/2}\sum_l\frac{(\pm1)^{l-1}}{l^{5/2}}\exp(l\mu/kT), \qquad \ldots\ldots(8.21)$$

$$\frac{N}{V}=n=\frac{1}{\lambda^3}\sum_l\frac{(\pm1)^{l-1}}{l^{3/2}}(\lambda^3z)^l=\left(\frac{2\pi mkT}{h^2}\right)^{3/2}\sum_l\frac{(\pm1)^{l-1}}{l^{3/2}}\exp(l\mu/kT), \qquad \ldots\ldots(8.22)$$

where use has been made of the relation (5.21): $z\lambda^3=\exp(\mu/kT)$ connecting z and μ.

These expressions are the well-known formulae for the pressure and energy of a Bose–Einstein or Fermi–Dirac gas, the last expression (8.22) determining the value of μ as a function of n and T; the upper sign is valid for Bose–Einstein and the lower sign for Fermi–Dirac statistics. *Thus the consequent treatment by quantum statistical mechanics of the ideal gas leads to the thermal and caloric equations of state of the ideal gas in Bose–Einstein or Fermi–Dirac statistics, depending on whether the eigenfunctions of the gas are symmetrical or antisymmetrical in the eigenfunctions of the molecules* (Kahn 1938).

The contribution (8.18) may therefore be said to account for the effect of Bose–Einstein or Fermi–Dirac statistics on the second virial coefficient of the equation of state.

We will not go into a more detailed treatment of the deviations from the ideal gas laws which result from the symmetry properties of the wave functions and which are described by the formulae (8.20) and (8.21). As is well known the Bose–Einstein case leads to a condensation phenomenon when the temperature becomes so low and the density so high that the thermodynamic potential μ becomes zero; this phenomenon has given rise to much discussion (cf. Einstein 1924, 1925, Uhlenbeck 1927, Kahn 1939, Chap. II, § 5, London 1938), the more because essential parts were taken as a basis for London's theory of the strange behaviour of liquid helium II. These problems, which are of the same kind as those connected with the condensation phenomena in real gases, are not considered in this survey.

(v) *Quantum Theory of Real Gases at High Temperatures*

The *second virial coefficient* for the non-ideal gas now is in general

$$B(T) = -\frac{N}{2}\int [W(r_{12}) - W_{\mathrm{id}}(r_{12})]\, d\mathbf{r}_{12} + B_{\mathrm{id}}(T). \quad \ldots\ldots (8.23)$$

This expression will first be applied at those temperatures where the deviations from classical theory are so small that the series expansion into powers of h can be used; this was originally due to Kirkwood (1931) and was applied to this case by Uhlenbeck and Beth (1936). Substituting the series expansion for $W(r_{12})$ given in § 6 (iii) (6.29) for $N=2$ and carrying out the integration, neglecting the influence of symmetry effects on $W(r_{12})$, one obtains an expansion of the second virial coefficient into powers of h of the form

$$B(T) = B^{(0)}(T) + B_{\mathrm{id}}(T) + B^{(1)}(T) + B^{(2)}(T), \quad \ldots\ldots (8.24)$$

where $B^{(0)}(T)$ is the classical expression for B and where $B^{(1)}$ and $B^{(2)}$ are the contributions proportional to h^2 and h^4 to the second virial coefficient; these are integrals containing $\exp(-\beta\phi(r))$ and the first and higher derivations of $\phi(r)$ with respect to r in the integrand.

The expressions $B^{(1)}$ and $B^{(2)}$ have been evaluated by de Boer and Michels (1938), using the intermolecular potential field (8.12). The result is expressed in the form

$$B^{(1)*} = \frac{2\pi}{3}\Lambda^{*2}\left(\frac{4}{T^*}\right)^{13/12}\Sigma c_n^{(1)}\left(\frac{4}{T^*}\right)^{n/2}, \quad \ldots\ldots (8.25)$$

$$B^{(2)*} = \frac{2\pi}{3}\Lambda^{*4}\left(\frac{4}{T^*}\right)^{23/12}\Sigma c_n^{(2)}\left(\frac{4}{T^*}\right)^{n/2}. \quad \ldots\ldots (8.26)$$

For the values of the coefficients $c_n^{(1)}$ and $c_n^{(2)}$ the reader is referred to the original publication. The quantity Λ^* which appears in these expressions is defined as $\Lambda^* = h/\sigma(m\epsilon)^{\frac{1}{2}}$. The reduced value $B^* = B/N\sigma^3$ thus depends not only on the reduced temperature T^*, but also on the value of this quantum mechanical parameter, in agreement with the quantum version of the law of corresponding states. Evidently the contribution B_{id} can also be written in reduced form:

$$B^*_{\mathrm{id}} = \mp\frac{1}{16\pi^{3/2}}\frac{\Lambda^{*3}}{T^{*3/2}}. \quad \ldots\ldots (8.27)$$

Using these expressions for $B^{*(1)}$, $B^{(2)*}$, B^*_{id} and the classical expression $B^{(0)*}$, we are able to determine the value of ϵ and σ for those gases to which quantum

theory has to be applied. The procedure is one of successive approximation: first values of ϵ/k and $N\sigma^3$ are determined from the classical formula (8.13) and the experimental values of the second virial coefficient at high temperatures where the deviations from classical theory are small. From these values of ϵ/k and $N\sigma^3$, using m, the value of Λ^* can be calculated. Then using this preliminary value of Λ^*, the quantum corrections (8.25) and (8.26) can be calculated to a first approximation, and using these theoretical values for $B^*(T^*)$ improved values of ϵ/k and $N\sigma^3$ can be obtained. Then the procedure can be repeated until a consistent set of values of ϵ/k and $N\sigma^3$ is obtained. In this comparison the theoretical curve for $B(T)$ can only be obtained at those temperatures at which the series expansion into powers of h converges sufficiently; this limits the temperatures for helium to those above 40° K. and for hydrogen and deuterium to those above 75° K. and 45° K. respectively. The values so obtained for the constants $N\sigma^3$, ϵ/k, ϵ and σ are

Gas	ϵ/k (° K.)	$N\sigma^3$ (cm^3)	ϵ (10^{-16}ergs)	σ (A.)	Λ^*
Helium	10·22	10·06	14·13	2·56	2·67
Hydrogen	37·0	15·12	51·10	2·92	1·73
Deuterium	37·0	15·12	51·10	2·92	1·22

In Figure 5 the theoretical curves are compared with the experimental data reduced by using the values of ϵ/k and $N\sigma^3$ given in this table. The experimental

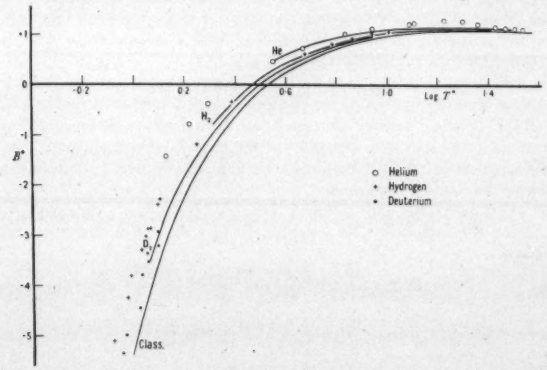


Figure 5. Quantum theoretical curves and experimental points for second virial coefficient B^* as function of log T^*.

data for hydrogen and deuterium are those measured by Schäfer (1937). These measurements are in general at too low temperatures to apply the series development into powers of h, but the extrapolated theoretical curves seem to fit the experimental values quite well. In Figure 6 the theoretical curve is compared with the small difference in B-value of H_2 and D_2 at room temperatures according to Michels and Goudeket (1941).

In a similar way the effect of quantum theory on the value of the *Joule–Thomson coefficient* μ_0 can be calculated using the formula (8.9). The resulting curves for helium and hydrogen have been drawn in Figure 7. A similar comparison has been made by Hirschfelder and his collaborators (1938).

(vi) *Quantum Theory of the Real Gas at very Low Temperatures*

At low temperatures the series expansion into powers of h no longer converges and new methods have to be developed for the evaluation of the second virial coefficient. Here the only measurements are those of helium, which have been carried out even down to temperatures of about 2° K. It is of advantage to

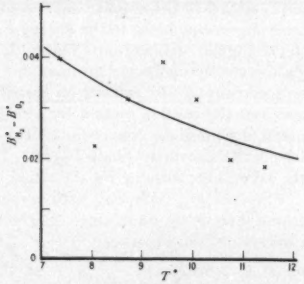


Figure 6. Quantum theoretical curve and experimental values of Michels and Goudeket for difference of second virial coefficient B^* of hydrogen and deuterium, between $T^*=7$ and 12 (0°−150° c.)

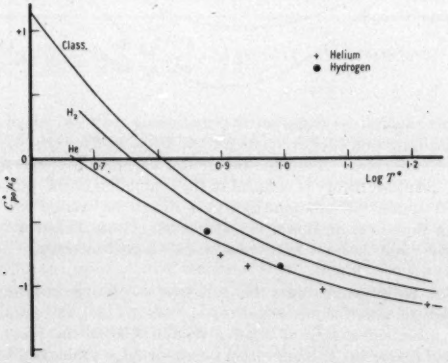


Figure 7. Quantum theoretical curve and experimental points for Joule-Thomson coefficient as function of log T^*.

transform the expression (8.23) for $B(T)-B_{\mathrm{id}}(T)$ in the following way: For $W(r_{12})$ is substituted the expression using energy eigenfunctions

$$W(r_{12})=\lambda^6 2!\,\Sigma|\psi_\sigma(\mathbf{r}_1,\mathbf{r}_2)|^2\exp(-\beta E_\sigma).$$

The integration over $\mathbf{r}_1$ and $\mathbf{r}_2$ is carried out, giving

$$B(T)=B_{\mathrm{id}}(T)-\frac{N}{2V}\lambda^6 2!\,(\Sigma_\sigma\exp(-\beta E_\sigma)-\Sigma_\sigma\exp(-\beta E_{\mathrm{id}\sigma})) \quad \ldots\ldots(8.28)$$

where E_σ and $E_{\mathrm{id}\sigma}$ are the eigenvalues of a pair of interacting and a pair of non-interacting molecules corresponding to the set of quantum numbers σ. If now in the quantum mechanical treatment of this two molecule problem, relative coordinates and coordinates of the centre of gravity are introduced, the motion of the latter can be described by classical mechanics and the integration over these coordinates can be carried out, resulting in a factor $2^{3/2}\lambda^{-3/2}V$. The states of relative motion are then specified by the angular momentum quantum number l and the quantum number k corresponding to the energy $E_k=\hbar^2k^2/m$, each state corresponding to a definite angular momentum l being $(2l+1)$-fold degenerate. The difference between the two terms in the summation over k depends in the continuous part of the spectrum ($E>0$) entirely on the difference between the density of quantum states of the relative motion for two interacting molecules and that of two non-interacting (ideal gas) molecules. This difference of density of states can be expressed in the so-called "phase" $\eta_l(k)$. The phase η is defined as the phase shift of the asymptotic formula for the radial wave function of two interacting molecules: $\psi_{kl}\sim\sin(kr_{12}-\frac{1}{2}l\pi+\eta)$, with respect to the asymptotic wave function of two non-interacting molecules: $\psi_{\mathrm{id}kl}\sim\sin(kr_{12}-\frac{1}{2}l\pi)$. The wave function ψ_{kl} is a solution of the equation

$$-\frac{\hbar^2}{m}\frac{d^2r\psi}{dr^2}+\left(\frac{\hbar^2}{m}\frac{l(l+1)}{r^2}+\phi(r)\right)r\psi=\frac{\hbar^2k^2}{m}r\psi \qquad \ldots\ldots(8.29)$$

and $\psi_{\mathrm{id}kl}$ is the solution of the corresponding equation in which $\phi(r)=0$. The resulting expression for B derived by Beth and Uhlenbeck (1937) and by Gropper (1937) then finally reads

$$B(T)=B_{\mathrm{id}}(T)-N2^{3/2}\lambda^{3/2}\sum_l(2l+1)\left\{\int\exp\left(-\frac{\hbar^2k^2}{mkT}\right)\frac{1}{\pi}\frac{d\eta}{dk}\,dk+\sum_{\mathrm{discr.}}\exp\left(-\frac{E}{kT}\right)\right\} \qquad \ldots\ldots(8.30)$$

where the first integral results from the continuous part ($E>0$) of the energy spectrum and the second sum accounts for the discrete states ($E<0$) of an interacting pair of molecules. Thus the theoretical calculation of the second virial coefficient in quantum theory is reduced to the evaluation of the phases $\eta_l(k)$ and of the discrete states. The summation over l must be carried out either only over the even values (symmetrical eigenfunctions; Bose–Einstein statistics) or only over the odd values (antisymmetrical eigenfunctions; Fermi–Dirac statistics).

Only at low temperatures does this sum over l converge sufficiently well to make the series suitable for numerical calculation. This can be understood as follows: for a certain energy of relative motion $\hbar^2k^2/m$, the phase shift $\eta_l(k)$ decreases when l increases and approaches zero for large values of l because then the repulsive term proportional to $l(l+1)/r^2$ always dominates the potential field $\phi(r)$: the influence of $\phi(r)$ disappears and just this influence makes the phases different from zero. The lower the value of k, the lower also the value of l for which the phases are negligible. As the values of k which are of importance are determined completely by the temperature T, it is clear that only at low temperatures the number of values of l, for which the phases are different from zero, is small enough to make numerical work possible.

Numerical calculations of the phases $\eta_l(k)$ and of the second virial coefficient of helium have been carried out by Massey and Buckingham (1938, 1939), by

de Boer and Michels (1939) and by Buckingham, Hamilton and Massey (1941) using different types of potential fields. Massey and Buckingham used a potential field 1·3 times as large as that calculated theoretically by Slater (1928), the field used being of the form $b\exp(-r/a)-c/r^6$, with $b=10^{-9}$ ergs, $c=1{\cdot}91\times10^{-12}$ ergs A. and $a=0{\cdot}217$ A. The phases were calculated by integration of the wave equation

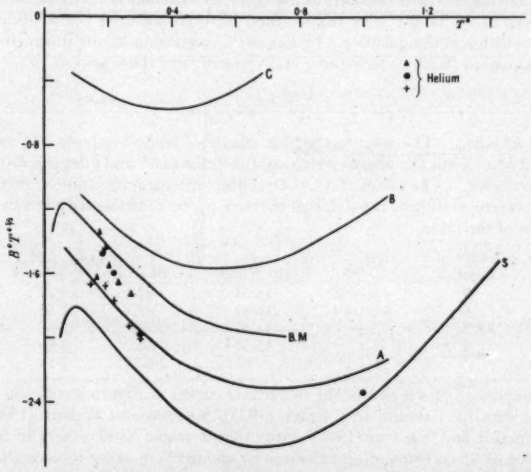


Figure 8. Quantum theoretical curves for $B^*T^{*0.2}$ of helium at very low temperatures (S, A, B.M, B, C) using different potential fields ($T^*=1$ corresponds to 10·2° K.). Experimental points: + Keesom and Walstra (1939), ● Keesom and Kraak (1935), ▲ Kistemaker and Keesom (1946).

of the relative motion of two molecules using a model differential analyser. From the asympotic behaviour of the radial wave function $\psi_{kl}(r)$, the phases were obtained for values of l ranging up to 10 and for each value of l for a suitable range of values of the energy of relative motion $\hbar^2k^2/m$. These results allowed calculations to be made up to temperatures of 25° K. A stationary state appears at about -0.0058ϵ being the total depth of the potential field, which contributes 12% to the second virial coefficient at 1° K. but only about 1·3% at 7° K.! These results are given in Figure 8 (S). De Boer and Michels (1939) calculated the phases up to an angular momentum $l=\sigma$ for the potential field $4\epsilon[(\sigma/r)^{12}-(\sigma/r)^6]$, with the values of the constants ϵ and σ determined from the second virial coefficient at temperatures above 40° K., by numerical integration of the wave equation (8.29). The results could be used to calculate the second virial coefficient of helium up to 5° K. and the curve is given in Figure 8 (B.M). Finally, Buckingham, Hamilton and Massey (1941) calculated the second virial coefficient with the same method for a number of potential fields of the form: $b\exp(-ar)-cr^{-6}-dr^{-8}$. In curves B and C all constants b, c and d are multiplied by a factor 0·8 and 0·6, keeping a constant. This leaves the minimum at the same place and only changes

the depth. In the case of curve A a stationary state appears at about 0.0007ϵ where ϵ is the total depth of the potential field. The contribution of this stationary state to the second virial coefficient amounts to about 1% at 1° K. and 0·1% at 7° K. To compare the potential fields of the five curves drawn in the figure, for each of these curves a potential field of the form $\phi(r)=4\epsilon[(\sigma/r)^{12}-(\sigma/r)^{6}]$ has been chosen which fits the real field as closely as possible. The values of the constants ϵ and σ are given in the table. The real quantity of importance for the solution of the wave equation is the quantity $\Lambda^*=h/\sigma(m\epsilon)^{\frac{1}{2}}$: expressing all the quantities of the wave equation (8.29) in terms of σ and ϵ one obtains the equation

$$-\frac{d^2r^*\psi}{dr^{*2}}+\frac{l(l+1)r^*\psi}{r^{*2}}+\frac{16\pi^2}{\Lambda^*}\left(\frac{1}{r^{*12}}-\frac{1}{r^{*6}}\right)r^*\psi=k^{*2}r^*\psi \quad \ldots\ldots(8.31)$$

where $k^*=k\sigma_0$. The solutions of this equation depend only on the values of Λ^* and k^*. Thus the phases which are functions of k^* and l depend also on Λ^* as a parameter. The values of Λ^*, the quantum mechanical parameter introduced above, corresponding to the different effective potential fields, are given in the last column of the table.

Curve of Figure 8	E^*_{stat}	ϵ (10^{-16} ergs)	σ (A.)	$\Lambda^*=\frac{h}{\sigma(m\epsilon)^{\frac{1}{2}}}$
S	−0·0058	16·34	2·62	2·42
A	−0·0004	14·69	2·62	2·56
B.M	0	14·13	2·56	2·67
B	—	12·24	2·62	2·80
C	—	9·80	2·62	3·14

Comparison of the plot of the theoretical curves in Figure 8 with the experimental data of Keesom and Kraak (1935), Keesom and Walstra (1939) and Kistenmaker and Keesom (1946) shows that it might be necessary to increase the value of Λ^* corresponding to helium by about 5% in order to make the curve lie between the curves B.M and A in Figure 8.

§9. CALCULATION OF THE MOLECULAR PAIR DISTRIBUTION FUNCTION

(i) *Series Development into Powers of the Density*

The formal expressions given in §7 for the series development of $n_2(r_{12})$ into powers of the density will be evaluated in this paragraph, as has been done with the equation of state in §8, by substituting the same expression

$$\phi(r)=4\epsilon[(\sigma/r)^{12}-(\sigma/r)^{6}]$$

for the intermolecular potential field with characteristic constants ϵ and σ.

In *classical statistics* the expression for $n_2(r_{12})$ which, after substituting (7.13) in (7.25), reads:

$$n_2(r_{12})=n^2W_2(r_{12})+n^3\int[W_3(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)-W_2(r_{12})W_2(r_{13})-W_2(r_{12})W_2(r_{23})+W_2(r_{12})]d\mathbf{r}_3+O(n^4)\ldots. \quad \ldots\ldots(9.1)$$

can be transformed by using the asumption of the additivity of intermolecular forces into

$$n_2(r_{12})=n^2\exp(-\beta\phi(r_{12})\cdot\{1+n\int(\exp\{-\beta\phi(r_{13})\}-1)\times(\exp\{-\beta\phi(r_{23})\}-1)d\mathbf{r}_3+O(n^2)\ldots.\}. \quad \ldots\ldots(9.2)$$

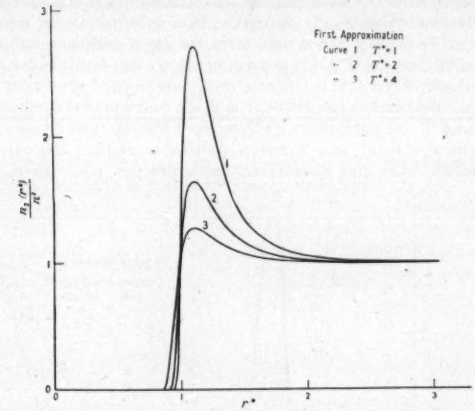


Figure 9. Classical distribution function in first approximation: $n_2(r^*)/n^2=\exp\{-\beta\phi(r^*)\}$.

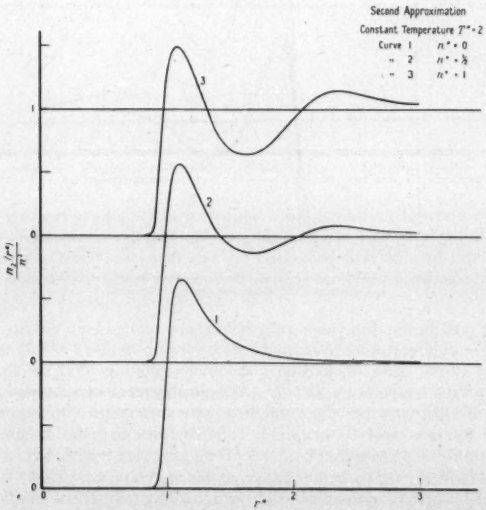


Figure 10. Classical distribution function in second approximation (equation(9.2)). Different densities $n^*=n\sigma^3$. Constant temperature T^*.

Thus at low densities the molecular pair distribution function simply approaches to the Boltzmann factor $\exp\{-\beta\phi(r_{12})\}$. After substituting the potential field $\phi(r)=4\epsilon[(r/\sigma)^{-12}-(r/\sigma)^{-6}]$ the second term has been evaluated numerically by de Boer and Michels (1937, 1939) and, using a method of Laplace transformation, by Montroll and Mayer (1941). These results are in good agreement. Clearly, this molecular distribution function $n_2(r_{12})/n^2$ is a function of the reduced distance $r_{12}^*=r_{12}/\sigma$, and of the parameters: the reduced temperature $T^*=kT/\epsilon$ and the reduced density $n^*=n\sigma^3$; it is therefore possible to evaluate this expression for all gases which follow the same molecular interaction law. In Figure 9 the

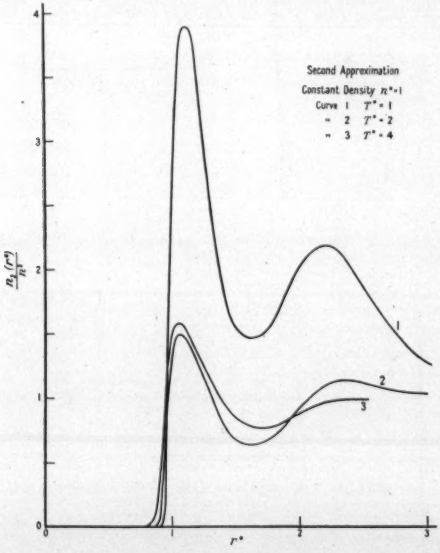


Figure 11. Classical distribution function in second approximation (equation (9.2)). Different temperatures T^*, constant density $n^*=n\sigma^3=1$.

molecular distribution function $n_2(r_{12}^*)/n^2$ has been plotted for a density approaching zero for different reduced temperatures, varying from $T^*=1$ up to $T^*=10$. In Figure 10 the effect of density is shown by plotting $n^2(r_{12}^*)/n^2$ for different densities n^* at a temperature $T^*=2$. At higher densities the second maximum appears characteristic for the molecular pair distribution function obtained from the Röntgen analysis of liquids (cf. § 10), although the densities are still low (the critical density is about $n^*=0{\cdot}8$), compared with liquid densities. In Figure 11 finally the variation of the maxima and mimima with T^* is shown by drawing the distribution function $n^2(r_{12}^*)/n^2$ according to (9.2) at a constant density $n^*=1{\cdot}0$ and for different temperatures. In sub-section (ii) the main features of these curves will be interpreted in terms of the potential of the average force.

In *quantum theory* the molecular distribution function differs from the value n^2 even in the case where *no interaction* between the molecules exists, because of the symmetry properties of the wave functions. At low densities, where only bimolecular encounters are of real interest, the molecular distribution function tends to

$$n_2(r_{12}) = n^2\left(1 \pm \exp - \frac{2\pi r_{12}^2}{\lambda^2}\right), \qquad \ldots\ldots(9.3)$$

where $\lambda = h/(2\pi mkT)^{\frac{1}{2}}$. In Figure 12 this function is given for symmetrical wave-functions (Bose–Einstein statistics) for helium molecules.

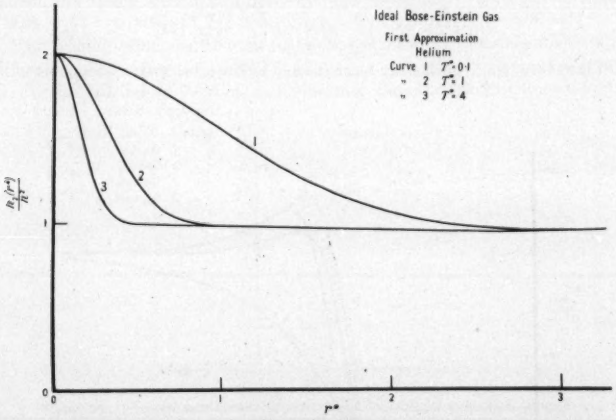


Figure 12. Quantum mechanical distribution function for ideal gas with molecules of mass of helium $T^*=1$ corresponds to 10·2° K., $r^*=1$ corresponds to 2·56 A. (equation (9.3)).

For *a real gas* of interacting molecules the distribution function in quantum theory can be evaluated relatively easily at high temperatures, where the deviations from quantum theory are small and the Kirkwood development into powers of h^2/mkT can be used. Here the expression is given only for the first term in the development (9.2), omitting symmetry effects,

$$\frac{n_2(r_{12})}{n^2} = \exp\{-\beta\phi(r_{12})\}.\left\{1 - \frac{\hbar^2}{6mkT}\left(\Delta\phi - \frac{1}{2kT}(\nabla\phi)^2\right) + \ldots.\right\}. \qquad \ldots\ldots(9.4)$$

The integrals for the next approximation containing $W(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ are still more difficult to evaluate than the corresponding classical ones. This first approximation shows already the typical effect of the wave character of the molecules: the plot of this function in Figure 13 for helium gas at reduced temperatures down to $T^*=4$ (40° K.) shows that the maximum is shifted to larger distances, the larger the de Broglie wavelength of the molecules. At very low temperatures $W(r_{12})$ tends simply to $|\psi_0(r_{12})|^2$; $\psi_0(r_{12})$ is the wave function corresponding to $E=0$, and use is made of the fact (§ 8 (iii)) that for helium no stationary states

exist. This limiting value of $W(r_{12})$ has been plotted in the same Figure. This shift of all mutual distances of the molecules to larger values can also be seen very clearly by comparing the experimental curves for liquid argon and helium in § 10.

(ii) *Distribution Function and Potential of Average Force*

Kirkwood (1935) has called attention again to the close relation which exists between the distribution function and the potential of the average force (cf. Onsager 1933). According to (5.12) the general expression for the molecular distribution function for a group of h molecules is

$$n_h(\mathbf{r}^h) = \frac{1}{(N-h)!\ Q_N} \int \exp\left(-\frac{\Phi(\mathbf{r}^N)}{kT}\right) d\mathbf{r}^{N-h}. \qquad \ldots\ldots(9.5)$$

On the other hand, the average force exerted by the other molecules on one of the

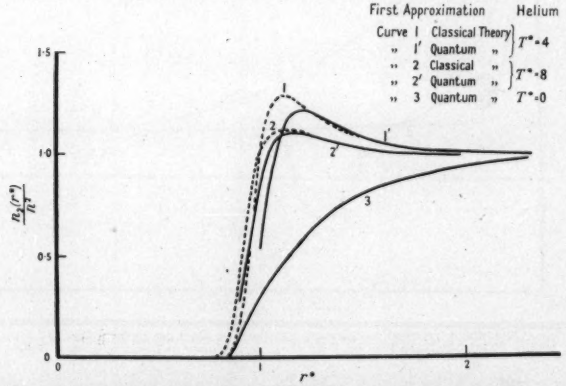


Figure 13. Quantum mechanical distribution for helium gas (equation (9.4)).

molecules ι of the group of h molecules, averaged over all positions of the remaining $N-h$ molecules and *keeping the group* $\{h\}$ *fixed in its configuration, is*

$$\overline{\mathbf{F}}_\iota(\mathbf{r}^h) = \frac{\int -\frac{\partial \Phi}{\delta \mathbf{r}_\iota} \exp\left(-\frac{\Phi(\mathbf{r}^N)}{kT}\right) d\mathbf{r}^{N-h}}{\int \exp\left(-\frac{\Phi(\mathbf{r}^N)}{kT}\right) d\mathbf{r}^{N-h}}. \qquad \ldots\ldots(9.6)$$

This average force on the molecule ι depends of course on the configuration $\mathbf{r}^h$ of the molecules of the group $\{h\}$. Evidently these two expressions are related to each other by the formula

$$\overline{\mathbf{F}}_\iota(\mathbf{r}^h) = kT \frac{\partial \ln n_h(\mathbf{r}^h)}{\delta \mathbf{r}_\iota}. \qquad \ldots\ldots(9.7)$$

It is useful to introduce here the concept potential of average force $\Psi(\mathbf{r}^h)$, according to

$$\bar{\mathbf{F}}_i(\mathbf{r}^h) = -\frac{\partial \Psi(\mathbf{r}^h)}{\delta \mathbf{r}_i}, \qquad n_h(\mathbf{r}^h) = z^h \exp\left(-\frac{\Psi(\mathbf{r}^h)}{kT}\right). \quad \ldots\ldots (9.8\,a, b)$$

The factor $Q_{N-h}/Q_N = z^h$ normalizes the potential of the average force so that

$$\exp\left(-\frac{\Psi(\mathbf{r}^h)}{kT}\right) = \frac{1}{N-h!\,Q_{N-h}} \int \exp\left(-\frac{\Phi(\mathbf{r}^N)}{kT}\right) d\mathbf{r}^{N-h}. \quad \ldots\ldots (9.9)$$

When the molecules of the set $\{h\}$ have no interaction with the other $N-h$ molecules Ψ should be equal simply to $\Phi(\mathbf{r}^h)$, but when the interaction is taken into account, deviations occur which have been developed in §7 into a power series of the activity z (equation (7.15), according to which $\exp -\beta\Psi(\mathbf{r}^h) = \Sigma l\, b_l^{(h)}(\mathbf{r}^h)\, z^{l-1}$.

The relation of the distribution function $n_2(r_{12})$ through the potential $\Psi(r_{12})$ of the average force to this force $\bar{\mathbf{F}}_2(r_{12})$ leads to an interesting interpretation of the shape of this distribution function, as, for instance, the molecular distribution func-

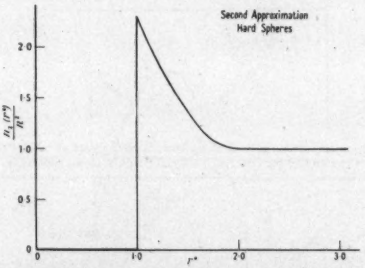


Figure 14. Classical distribution function in first approximation for hard spheres (equation (9.10)).

tion for *hard elastic spheres*. This can easily be calculated using the general series development (9.2) and substituting $\phi(r^*) = \infty$ for $r^* < 1$ and $\phi(r^*) = 0$ for $r^* > 1$. Then $n_2(r_{12}^*)$ becomes, taking only the terms with n^3,

$$\left.\begin{aligned} &n_2(r_{12}^*) = 0 && (r_{12}^* < 1), \\ &\frac{(n_2 r_{12}^*)}{n^2} = 1 + n^* \frac{4\pi}{3}\left(1 - \frac{3}{4} r_{12}^* + \frac{1}{16} r_{12}^{*3}\right) && (1 < r_{12}^* < 2), \\ &\frac{(n_2 r_{12}^*)}{n^2} = 1 && (2 < r_{12}^*). \end{aligned}\right\} \quad \ldots\ldots (9.10)$$

The plot of this function in Figure 14 shows that, even for hard spheres, the molecular pair distribution function differs from the value unity at distances larger than the diameter σ. In terms of the potential of the intermolecular force this can easily be understood. When the distance becomes smaller than 2σ, $n_2(r_{12})$ increases, so $\Psi(r_{12}^*) = -n\sigma^3(4\pi/3 - \pi r_{12}^* + \pi r_{12}^{*3}/12)kT$ decreases, giving rise to an *attractive average force* $F_2(r_{12}) = -n\pi\sigma^2(1 - r^{*2}/4)kT$. This average force, which acts on a molecule 2 in the direction of 1, can be easily calculated to

result from the shielding effect of the molecule 1 on the collisions of the remaining $(N-2)$ molecules with the molecule 2. This results in a force in the direction of molecule 1 as soon as the distance is smaller than 2σ. The space where a third molecule can have its centre is outside the curve drawn in Figure 15 (*a*).

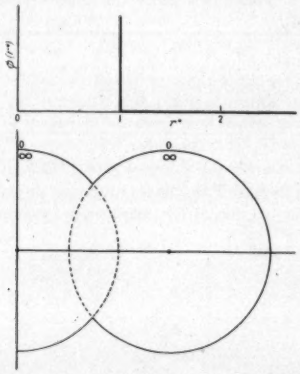


Figure 15 (*a*). The potential field surrounding two hard spherical molecules, showing region excluded for centre of a third neighbouring molecule.

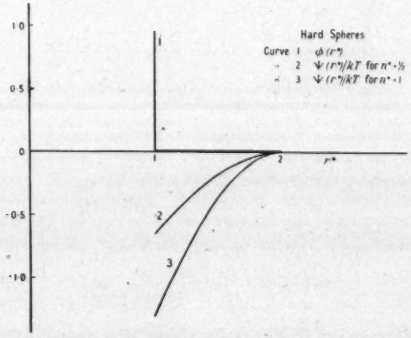


Figure 15 (*b*). Potential of average force for hard spheres $\psi(r^*)/kT$ for $n^*=n\sigma^3=1$ and $n^*=\frac{1}{2}$.

The potential of the average force for hard spheres has been plotted in Figure 15 (*b*).

A similar interpretation can be given of the distribution function of two molecules which interact with the *real potential field* $\phi(r)=4\epsilon[(\sigma/r)^{12}-(\sigma/r)^6]$. At low densities $\Psi(r_{12})$ becomes equal to $\phi(r_{12})$) but at higher densities deviations

occur; these are plotted in Figures 10 and 11. In the potential $\Psi(r_{12})$ these deviations appear in the form

$$\Psi(r_{12}) = \phi(r_{12}) - nkT\int(\exp\{-\beta\phi(r_{13})\}-1)(\exp\{-\beta\phi(r_{23})\}-1)\,d\mathbf{r}_3 + O(n^2)\ldots. \quad \ldots\ldots(9.11)$$

The first correction proportional to n accounts in the first approximation for the effect of the other molecules on the molecular distribution function. The second term on the right-hand side has been plotted (curves 2 and 3) for a temperature T^* equal to 1 and 2 and a density $n^* = 1{\cdot}0$ in Figure 16 (*a*). Thus the average force between the molecules 1 and 2 consists of the direct force $-d\phi(r_{12})/dr_{12}$ and the average force exerted by the other molecules, the potential of this contribution being curves 2 and 3. There thus occurs an extra repulsive average force at distances smaller than $r^* = 1{\cdot}5$, an extra attractive force between $r^* = 1{\cdot}5$ and $2{\cdot}2$ and, for still larger distances, an extra average force which is again repulsive. In Figure 16 (*b*) the potential field resulting from the two molecules 1 and 2 is drawn

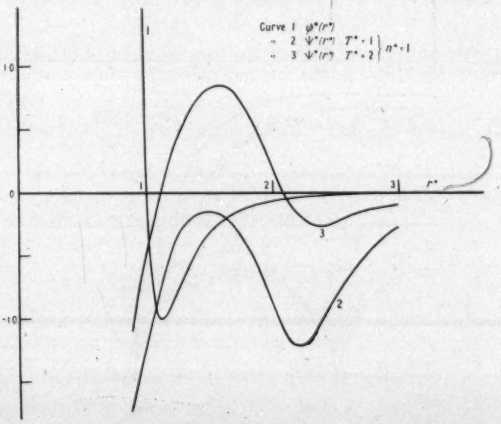


Figure 16 (*a*). Contributions to the potential of average force for real molecules: curve 1 ($\phi(r_{12})$) and curves 2 and 3 (second term in (9.11)).

at distances 1·3, 2·0 and 2·5 in the first, second and third region respectively. The extra average repulsive force in the first region results from the same cause as in the case of hard spheres: in the molecule 2 more collisions occur from the right than from the left, part of the space being occupied on the left by the molecule 1, so that a resultant force to the left results. In the second region ($r^* \approx 2{\cdot}0$) the deep potential minimum favours positions of a third molecule on the left of molecule 2 to such an extent that the extra repulsive force resulting from these molecules surpasses the attractive force resulting from collisions on the right-hand side of molecule 2. In the third region ($r^* \approx 2{\cdot}5$) the molecules between 1 and 2 at $r^* = 1{\cdot}25$ exert an attractive direct force on 1 and 2, and no screening effect of the molecule 1 exists, thus making again a resultant attractive average force between

the molecules 1 and 2. In this way the concept of the potential of the average force allows one to interpret physically the behaviour of the molecular distribution function.

(iii) *Integral Equation for* $n_2(r_{12})$

This concept of the potential of the average force allows one to derive the following integral equation for $n_2(r_{12})$. On one hand the average force $\overline{\mathbf{F}}_2(r_{12})$ acting on the molecule 2 of a pair of molecules 1 and 2 is related to $n_2(r_{12})$ by the

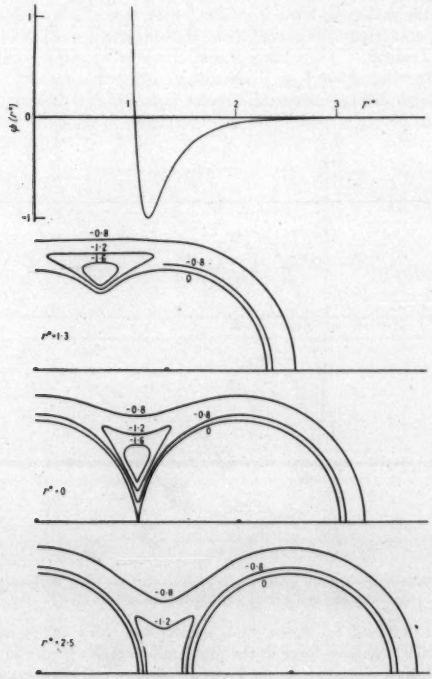


Figure 16 (*b*). The potential field surrounding two real molecules showing potential energy of molecule in units ϵ.

relation (9.7). On the other hand the average force can be calculated from (9.6) using the definition of the molecular distribution function $n_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ and $n_2(r_{12})$ according to (9.5), giving

$$kT\frac{\partial \ln n_2(r_{12})}{\partial \mathbf{r}_2} = -\frac{\partial \phi(r_{12})}{\partial \mathbf{r}_2} - \int \frac{\partial \phi(r_{23})}{\partial \mathbf{r}_2}\frac{n_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)}{n_2(r_{12})}\,d\mathbf{r}_3. \quad \ldots\ldots(9.12)$$

This integral equation is quite general except that in its derivation *classical statistics* have been used. The first term on the right-hand side is the direct force acting on the molecule 2 as a result of the presence of the molecule 1; the integral on the right-hand side gives the contribution of the other molecules to this average force, $n_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)/n_2(r_{12})$ being the probability of finding a third molecule in a position $\mathbf{r}_3$, the molecules 1 and 2 being kept fixed.

The equation (9.12) for n_2 can easily be generalized for higher distribution functions by equating (9.6) and (9.7) giving

$$kT\frac{\partial n_h(\mathbf{r}^h)}{\partial \mathbf{r}_i} = -\sum_{\kappa=1}^{h}\frac{\partial \phi(r_{i\kappa})}{\partial \mathbf{r}_i} n_h(\mathbf{r}^h) - \int \frac{\partial \phi(r_{i,h+1})}{\partial \mathbf{r}_i} n_{h+1}(\mathbf{r}^{h+1})\, d\mathbf{r}_{h+1}. \quad \ldots\ldots(9.13)$$

This gives a set of equations for $n_h(\mathbf{r}^h)$, each of which can be solved when $n_{h+1}(\mathbf{r}^{h+1})$ is known. So one has to make certain assumptions about $n_{h+1}(\mathbf{r}^{h+1})$ to get an equation for $n_h(\mathbf{r}^h)$.

Kirkwood (1935) approximated the integral expression on the right-hand side of (9.13) for $h=2$, by writing for $n_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)/n_2(r_{12})$ the Boltzmann factor

$$n\exp\{-\beta[\phi(r_{13})+\phi(r_{23})]\},$$

which will certainly be valid at not too high densities. Integration of the equation and determination of the integration constant so that $n_2(r_{12})$ approaches n^2 for $r_{12}\to\infty$ gives

$$n_2(r_{12}) = n^2\exp\left[-\frac{\phi(r_{12})}{kT} + n\int(\exp\{-\beta\phi(r_{13})\}-1)(\exp\{-\beta\phi(r_{23})\}-1)\,d\mathbf{r}_3\right]. \quad \ldots\ldots(9.14)$$

This equation gives up to powers of n^3 the same result as the equation (9.2) which is the exact power series into powers of n.

This assumption of Kirkwood can be generalized as follows:

$$n_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \frac{n_2(r_{12})n_2(r_{13})n_2(r_{23})}{n^3}. \quad \ldots\ldots(9.15)$$

This superposition assumption, whose merit nobody has assessed, but which might be assumed to be just somewhat better than writing

$$n_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = n_2(r_{12})\, n\exp -\beta[\phi(r_{13})+\phi(r_{23})],$$

leads to an integral equation for $n_2(r_{12})$ of the form

$$kT\frac{\partial \ln n_2(r_{12})}{\partial \mathbf{r}_1} = -\frac{\partial\phi(r_{12})}{\partial \mathbf{r}_1} - \frac{1}{n^3}\int\frac{\partial\phi(r_{13})}{\partial \mathbf{r}_1} n_2(r_{13})\, n_2(r_{23})\, d\mathbf{r}_3. \quad \ldots\ldots(9.16)$$

As has been shown by Born and Green (1946) this equation can be integrated over $\mathbf{r}_1$, giving, with $r_{13}=t$ and $r_{23}=s$,

$$\log\frac{n_2(r_{12})}{n^2} = -\frac{\phi(r_{12})}{kT} + n\int\int(s^2-t^2)\frac{t+r_{12}}{r_{12}}\frac{d\phi(s)}{ds}\left(\frac{n_2(t+r_{12})}{n^2}-1\right)\frac{n_2(s)}{n^2}\pi\, dt\, ds. \quad \ldots\ldots(9.17)$$

Green (1947) had to introduce some approximations to integrate this equation for $n_2(r_{12})$, the consequence of which cannot easily be seen. The result is that

$$\frac{n_2(r_{12})}{n^2} = \exp\{-\beta\phi(r_{12})\}\cdot\left\{1+\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{+\infty}\frac{[\epsilon sh(s)]^2}{v-\epsilon h(s)}\frac{\sin sr_{12}}{sr_{12}}ds\right\}, \quad \ldots\ldots(9.18)$$

where $v^{-1}=(2\pi)^{3/2}n$ and where $sh(s)=(2\pi)^{-1}\int r_{12}[\exp\{-\beta\phi(r_{12})\}-1]\sin sr_{12}\,dr_{12}$, and where ϵ a kind of average value of $n_2(r_{12})/(n^2\exp-\beta\phi(r_{12}))$ for small r_{12}. Green then develops the integral occurring in (9.18) into powers of the density by assuming $\epsilon h(s)v^{-1}=n(2\pi)^{3/2}\epsilon h(s)$ to be small and developing the denominator. Substitution of the expression so obtained for $n_2(r_{12})$ into the equation (5.11) for the pressure then leads to the following development into powers of n:

$$pV=RT\left\{1-\Sigma\frac{k}{k+1}\alpha_k n^k\right\}, \qquad \ldots\ldots(9.19)$$

where $\alpha_1=(2\pi)^{3/2}h(0)$, $\alpha_k=\dfrac{(2\pi)^{(3k-1)/2}\epsilon^k}{2}\int_{-\infty}^{+\infty}[h(s)]^{k+1}s^2\,ds.$ $\ldots\ldots(9.20)$

This equation of state can be calculated relatively easily with the help of the equation for $h(s)$. It can be shown that when ϵ is taken to be unity the α_1 and α_2 become identical with the exact expressions β_1 and β_2 obtained with the application of the principle of Ursell's method in §7, equation (7.22).

The integral equation (9.16) for $n_2(r_{12})$ has been used by Kirkwood (1939) and Kirkwood and Boggs (1942) to obtain the molecular pair distribution function for hard spherical molecules and their results showed remarkable agreement with thx experimental curves for liquid argon.

§10. DISTRIBUTION FUNCTION: FLUCTUATIONS: SCATTERING OF LIGHT AND X-RAYS

(i) *Fluctuations*

The molecular pair distribution function is capable of direct experimental determination as a result of its intimate connection with the fluctuations in density in a gas or liquid system. This connection has been stressed by Ornstein and Zernike (1914) in a discussion of the critical opalescence phenomena, which are caused by the density fluctuations near the critical point, and by Zernike and Prins (1927), who applied the theory to x-ray diffraction phenomena. An excellent survey of this field and important new contributions were given by Yvon (1937). The fundamental formula can easily be derived as follows. Introduce the δ-function $\delta(\mathbf{r}_\iota-\mathbf{r})$, where $\mathbf{r}_\iota$ is the coordinate at a certain time of the molecule ι. The number of molecules n_Ω contained in a certain volume Ω is then given by

$$n_\Omega=\sum_{\iota=1}^{N}\int_\Omega\delta(\mathbf{r}_\iota-\mathbf{r})\,d\mathbf{r}, \qquad \ldots\ldots(10.1)$$

similarly

$$\begin{aligned}n_\Omega^2&=\sum_{\iota=1}^{N}\sum_{\kappa=1}^{N}\int_\Omega\int_\Omega\delta(\mathbf{r}_\iota-\mathbf{r})\,\delta(\mathbf{r}_\kappa-\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}'\\&=\sum_{\iota=1}^{N}\int_\Omega\delta(\mathbf{r}_\iota-\mathbf{r})\,d\mathbf{r}+\sum_{\iota>\kappa=1}^{N}\sum_{}^{N}\int_\Omega\int_\Omega\delta(\mathbf{r}_\iota-\mathbf{r})\delta(\mathbf{r}_\kappa-\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}'. \qquad \ldots\ldots(10.2)\end{aligned}$$

The average values of n_Ω and n_Ω^2 can be expressed in terms of probability densities, and in this case the generic distribution functions lead immediately to

$$\bar{n}_\Omega=\int_\Omega n_1(\mathbf{r}_1)\,d\mathbf{r}_1, \qquad \ldots\ldots(10.3)$$

$$\overline{n_\Omega^2}=\int_\Omega n_1(\mathbf{r}_1)\,d\mathbf{r}_1+\int_\Omega\int_\Omega n_2(\mathbf{r}_1,\mathbf{r}_2)\,d\mathbf{r}_1\,d\mathbf{r}_2. \qquad \ldots\ldots(10.4)$$

Now the fluctuations of n_Ω are given by $\overline{(n_\Omega - \bar{n}_\Omega)^2} = \overline{n_\Omega^2} - \overline{n_\Omega}^2$. This gives

$$\overline{n_\Omega^2} - \overline{n_\Omega}^2 = \int_\Omega n_1(\mathbf{r}_1)\, d\mathbf{r}_1 + \int_\Omega \int_\Omega [n_2(\mathbf{r}_1, \mathbf{r}_2) - n_1(\mathbf{r}_1) n_1(\mathbf{r}_2)]\, d\mathbf{r}_1\, d\mathbf{r}_2. \quad \ldots\ldots(10.5)$$

Dividing this by $\bar{n}_\Omega$, one obtains the following expression for the fluctuation of n_Ω:

$$\frac{\overline{n_\Omega^2} - \overline{n_\Omega}^2}{\bar{n}_\Omega} = 1 + n \int_\Omega \left[\frac{n_2(r_{12})}{n^2} - 1 \right] d\mathbf{r}_{12}, \quad \ldots\ldots(10.6)$$

which relates the fluctuation of the number of molecules in Ω with the molecular pair distribution function $n_2(r_{12})$. The integration on the right-hand side could now be extended over the entire phase space if $n_2(r_{12})$ approached n^2 for $r_{12} \to \infty$, as then the integrand vanishes when r_{12} is large compared with the range of intermolecular forces. In fact, however, $n_2(r_{12})$ can be developed in a power series:

$$n_2(r_{12}) = n^2 \left[g(r_{12}) + \frac{h(r_{12})}{N} + O\left(\frac{1}{N^2}\right) \ldots \right]. \quad \ldots\ldots(10.7)$$

The terms of the order of $1/N$ have been neglected in § 7, but are here of importance since their contribution is not negligible when the integration is carried out over the entire phase space. In this formula $g(r_{12}) \to 0$ for $r_{12} \to \infty$, and $h(r_{12}) \to h$ for $r_{12} \to \infty$. Substitution of (10.7) in (10.6) gives

$$\frac{\overline{n_\Omega^2} - \overline{n_\Omega}^2}{\bar{n}_\Omega} = 1 + n \int_0^\infty [g(r_{12}) - 1]\, d\mathbf{r}_{12} + \frac{n_\Omega}{N} h, \quad \ldots\ldots(10.8)$$

where $n \int_\Omega h\, d\mathbf{r} \approx h n_\Omega$. It can now be shown, that this last term can be neglected when $n_\Omega \ll N$, by substituting in the relation

$$\int n_2(\mathbf{r}_{12})\, d\mathbf{r}_2 = (N-1) n_1,$$

which expression for $n_2(\mathbf{r}_{12})$ follows from the definition (equation (3.13)) of the generic distribution functions. This substitution leads to

$$1 + n \int_0^\infty [g(\mathbf{r}_{12}) - 1]\, d\mathbf{r}_{12} + h = 0. \quad \ldots\ldots(10.9)$$

Comparison of (10.9) and (10.8) shows that the term $n_\Omega h/N$ can indeed be neglected when $n_\Omega \ll N$. Thus:

$$\frac{\overline{n_\Omega^2} - \overline{n_\Omega}^2}{\bar{n}_\Omega} = 1 + n \int_0^\infty [g(r_{12}) - 1]\, dr_{12}. \quad \ldots\ldots(10.10)$$

This connects the fluctuations in density formula with the molecular pair correlation function $g(r_{12})$.

On the other hand, statistical thermodynamics (cf. Tolman 1938, *p.* 629, etc.) leads to an expression which connects the fluctuation of the density with macroscopic quantities:

$$\frac{\overline{n^2} - \overline{n_\Omega}^2}{\bar{n}} = -\frac{RT}{V} \left(\frac{\partial V}{\partial P} \right)_T;$$

this formula was derived by Gibbs in 1902. Indeed the substitution of the development (7.16) for

$$n_2(r_{12}) = \Sigma l b_l^{(2)}(r_{12}) z^{l+1}$$

in (10.10) and integration over r_{12}, using the definitions (7.6) and (7.13) for $b_l^{(2)}$ and b_l, leads to $(1/n)\Sigma l^2 b_l z^l$, which can be shown to be equal to (10.11), using the equation $p = kT\Sigma b_l z^l$ and $n = \Sigma l b_l z^l$ derived in § 7.

Because of this relation to the fluctuations in density, the pair correlation function $g(r_{12})$ is also intimately related to the phenomena of the scattering of visible light and of röntgen waves, which have led to the important method of derivation of the pair correlation-function in liquids and gases by x-ray analysis.

(ii) *Scattering of Electromagnetic Waves*

It is useful to review in short the method of calculating the scattered intensity of electromagnetic waves which traverse a medium consisting of molecules: gas or liquid. When no density fluctuations exist the waves scattered by the different molecules of the fluid cancel by interference, so that no resulting scattering of electromagnetic waves exists. The scattering results entirely from the density fluctuations in the gas or fluid (cf. Born 1933, § 81).

Consider the electromagnetic waves scattered by the molecules $\iota, \kappa, \lambda, \ldots$. of the fluid the first of which is supposed to be situated at the origin. If the wave vector of the incident wave is given by $\mathbf{k}_0$, where $k_0 = 2\pi/\lambda$ and the scattered waves in the direction of $\mathbf{k}$ are investigated ($k_0 = k$), the phase difference between the waves scattered by the two molecules ι and κ is $(\mathbf{k}_0 - \mathbf{k}) . \mathbf{r}_{\iota\kappa}$. If the incident electromagnetic waves are supposed to have the electric vector perpendicular to the plane of Figure 17, the electric field in a direction θ at a distance R, large

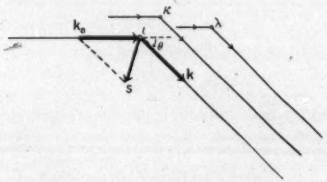


Figure 17. Incident and scattered rays.

compared with the dimension of the volume Ω in which the scattering fluid is contained, is given by

$$\mathbf{E} \propto \frac{1}{R} \Sigma_\kappa \mathbf{A}_\kappa \exp\{i(\mathbf{k}_0 - \mathbf{k}) . \mathbf{r}_{\iota\kappa}\}. \qquad \ldots\ldots (10.11)$$

The amplitude $\mathbf{A}_\kappa$ depends on the character of the electromagnetic waves, which are considered.

(*a*) For *light waves* $\mathbf{A}_\kappa = \boldsymbol{\mu}_\kappa k^2$, where $\boldsymbol{\mu}_\kappa$ is the dipole moment induced by the electromagnetic waves in the molecule κ. If the fluid is dilute the electric field acting on each molecule is practically equal to the electric field $\mathbf{E}_0$ of the incident radiation. When α is the polarizability of the molecules $\boldsymbol{\mu}_\kappa = \alpha \mathbf{E}_0$ and $\mathbf{A}_\kappa = \alpha k^2 \mathbf{E}_0$. For dense fluids the induced electric moments of the surrounding molecules also contribute to the effective field acting on a definite molecule. The problem of calculating the effective field in this case has been solved by Kirkwood (1936) and by Yvon (1937), who showed the close connection which this effective field has with the molecular distribution functions. This result will not be discussed

here, however, and for brevity the effective field is assumed to be equal to $\frac{1}{3}(\epsilon+2)\mathbf{E}_0$, which accounts to a first approximation for the increase of the effective field in dense media. This leads to the value $\mathbf{A}=\frac{1}{3}\alpha k^2(\epsilon+2)\mathbf{E}_0$.

(*b*) For *Röntgen waves* the value of $\mathbf{A}$ for a single electron is $(e^2/mc^2)\mathbf{E}_0$. An atom consisting of a number of Z electrons does not increase the scattering by a factor Z, because part of the scattering intensity is cancelled by interference of the scattered waves of the different electrons. Thus the total value of $\mathbf{A}$ for an atom of Z electrons is given by $\mathbf{A}=\mathbf{E}_0(e^2/mc^2)f(k,\theta)$, where f, the atomic scattering factor, depends on the scattering angle θ and has maximum value Z. When the fluid does not consist of monatomic molecules, the scattering intensity by a single molecule is more complicated by the interference of the waves of the atoms which constitute one molecule.

The intensity of the scattered radiation at a distance R from the scattering volume Ω in a direction θ is, substituting $\mathbf{s}=\mathbf{k}-\mathbf{k}_0$, obtained by taking the square of (10.11)

$$|\mathbf{E}|^2 \propto \frac{A^2}{R^2}\Sigma_\kappa\Sigma_\lambda \exp\{i\mathbf{s}\,.\,\mathbf{r}_{\kappa\lambda}\}, \qquad \ldots\ldots(10.12)$$

which expression must be averaged over all vectors $\mathbf{r}_{\kappa\lambda}=\mathbf{r}_\lambda-\mathbf{r}_\kappa$. The averaging procedure over all directions of $\mathbf{r}_{\kappa\lambda}$ with respect to $\mathbf{s}$ gives $\sin sr_{\kappa\lambda}/s_\kappa r_{\kappa\lambda}$ where s is the length of $\mathbf{s}$, given by $s=2k\sin\theta=4\pi\sin\theta/\lambda$. The summation over all κ and λ can now be expressed in terms of the generic distribution functions, which gives two terms, the first resulting from $\kappa=\lambda$ and the second from $\kappa\neq\lambda$ respectively. Thus:

$$I(\theta)=\frac{A^2}{R^2}\left[\int_\Omega n_1(\mathbf{r}_1)d\mathbf{r}_1+\int_\Omega\int_\Omega n_2(\mathbf{r}_1\mathbf{r}_2)\frac{\sin sr_{12}}{sr_{12}}\,d\mathbf{r}_1\,d\mathbf{r}_2\right]. \qquad \ldots\ldots(10.13)$$

Finally this expression can be transformed, using the formula

$$n_2(r_{12})=n_1^2[g(r_{12})+h(r_{12})/N+\ldots.]$$

introduced earlier (cf. Yvon 1937 b, Chap. V).

(*a*) For *light waves* the formula can be simplified since the wavelength λ is large. Then $(\sin sr_{12})/sr_{12}$ can be put equal to unity for all values of r_{12} for which $g(r_{12})$ differs from zero. Introduction of the value if $\mathbf{A}$ for light waves, viz. $\mathbf{A}=\alpha\frac{1}{3}(\epsilon+2)\mathbf{E}_0(2\pi/\lambda)^2$, then gives

$$I(\lambda)=\frac{16\pi^4\alpha^2(\epsilon+2)^2E_0^2n_\Omega}{9R^2\lambda^4}\left[1+n\int_0^\infty[g(r_{12})-1]d\mathbf{r}_{12}\right]. \qquad \ldots\ldots(10.14)$$

The scattered intensity is thus, using (10.10), proportional to the fluctuations in the number density $n_\Omega^2-\overline{n_\Omega}^2$. As these fluctuations can be expressed independently by statistical thermodynamics in thermodynamic quantities, the intensity of light scattering can be calculated without explicit knowledge of the molecular distribution function. For a detailed survey of the problems connected with (10.14), the reader is referred to the monograph by Yvon (1937).

(*b*) For *Röntgen waves* one obtains on substituting the correct value of $A=(e^2/mc^2)E_0f(s)$ the well-known expression

$$I(s)=\frac{(e^2/mc^2)^2E_0f(s)^2n_\Omega}{R^2}\left[1+n\int_0^\infty[g(r_{12})-1]\frac{\sin sr_{12}}{sr_{12}}\,d\mathbf{r}_{12}\right]. \qquad \ldots\ldots(10.15)$$

This formula derived by Zernike and Prins (1927) enables calculation of the intensity of x-ray scattering as function of $s=4\pi\sin\theta/\lambda$ when $g(r_{12})$ is known.

In § 10 (iii) it will be shown that conversely also from the intensity of X-ray scattering as function of s, the molecular pair distribution function $g(r_{12})$ can be calculated as function of r_{12}.

(iii) *Molecular Pair Distribution Function from X-ray Analysis*

The formula (10.15) connecting $I(s)$ with $g(r_{12})$ has been transformed by Debye and Menke (1931) so as to make it possible to calculate $g(r_{12})$ from the experimental curve for $I(s)$. To make this transformation clear it is of advantage to write $I_0 = (e^2/mc^2)^2 E_0^2 f(s)^2 n_Q/R^2$, where I_0 is the intensity of the scattering which would occur when no part of the scattered intensity is cancelled by interference of the waves scattered by different atoms. Write $i(s) = I(s)/I_0(s)$, where now $i(s)$ is the ratio by which the scattering is reduced as a result of interference effects. Then (10.14) can be written in the symmetrical form

$$s[i(s)-1] = 4\pi \int_0^\infty r_{12} n[g(r_{12})-1] \sin(sr_{12})\, dr_{12}. \quad \ldots\ldots (10.16)$$

By Fourier transform, this can be transformed into

$$r_{12} n[g(r_{12})-1] = \frac{1}{2\pi^2} \int_0^\infty s[i(s)-1] \sin(sr_{12})\, ds. \quad \ldots\ldots (10.17)$$

This important formula of Debye and Menke allows one to calculate $g(r_{12})$ from the experimental values of $i(s)$, i.e. from the measurements of the scattered intensity of Röntgen waves as a function of $s = 4\pi \sin\theta/\lambda$. In general, the difficulty of this procedure lies in the exact knowledge of $i(s)$, which requires accurate knowledge of the experimental values both of $I(s)$ and of $I_0(s)$. This latter quantity contains the atomic scattering factor $f(s)$, which in most cases is not very accurately known. Moreover, in the case of multi-atomic molecules, interference effects occur also between the waves scattered by the atoms of one molecule, which still further complicates the $I(s)$ curve.

An excellent survey of all the work done in this field has recently been given by Gingrich (1943); the reader is referred to this paper for more details.

As a typical example attention may be directed to the results obtained with argon by Keesom and de Smedt (1922, 1923) and extended over an appreciable range of temperatures by Eisenstein and Gingrich (1940, 1942). Besides the experiments on liquid helium of Keesom and Taconis (1937, 1938), these are the only experiments carried out on the important, and relatively easily interpreted, monatomic substances. Figure 18 shows the results for the $g(r^*)$ as a function of $r^* = r/\sigma$ for liquid argon along the saturation curve; the curves (1), (2), ... (5) correspond respectively to the temperatures 84·4° K. (0·8 atm., somewhat lower than melting point), 91·8° K. (1·8 atm.), 126·7° K. (18·3 atm.), 144·1° K. (37·7 atm.) and 149·3° K. (46·8 atm., about the critical point). The pronounced maximum at the lowest temperatures for $r \sim 3{\cdot}8$ A., with $g(r) \sim 3{\cdot}6$, gradually decreases in importance when the temperature increases, lying at about 4 A., with $g(r) \sim 1{\cdot}4$ at the critical temperature. The first minimum, rather pronounced at $r = 4{\cdot}6$ A., with $g(r) \sim 0{\cdot}5$ at the melting point also gradually becomes less important with rising temperature: at the critical temperature it occurs at about 5·5 A., with $g(r) \sim 0{\cdot}9$.

The gradual smearing out of sharp maxima and minima which accompanies the rise in temperature and the small decrease in density of the saturated liquid,

is in general agreement with the theoretical behaviour of the molecular distribution function for compressed gases, where it can be calculated exactly by theoretical means (cf. § 8). The value of σ corresponding to argon is 3·41 A., so that the first maximum occurs at $r^* \sim 1{\cdot}1$ at the melting point and at about 1·2 in the liquid state near the critical point; these figures can be compared directly with the abscissae in Figure 11.

The experiments of Keesom and Taconis on liquid He I and He II showed the appearance of a maximum at about 3·5–4 A. with $g(r) \sim 1{\cdot}2$ and a minimum at

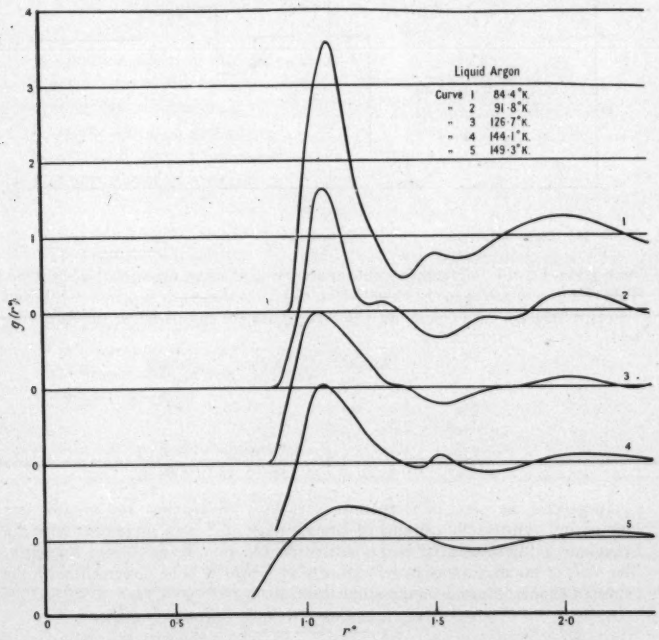


Figure 18. Experimental curves for $g(r^*)$ as function of $r^* = r/\sigma$ for liquid argon ($\sigma = 3{\cdot}42$ A.).

about 4·5–5 A. with $g(r) \sim 0{\cdot}8$. Remembering that the value of σ for helium is about 2·56 A. this maximum occurs at the value $r^* \sim 1{\cdot}5$, very much larger than that for argon. It is well known that this results from the influence of quantum mechanics, which is very large in the case of helium and which tends to increase the distance between the molecules, as has been discussed in § 7 (vi).

The only experiments on diatomic permanent gases which can be mentioned here are those on nitrogen, which were made originally by Keesom and de Smedt

(1922, 1923) and which were extended appreciably recently by Sharrah (1942). The results obtained by Sharrah at the triple points temperature 64° K. and at 89° K. in liquid nitrogen are complicated by the extra interference caused by the two atoms in the nitrogen molecule. The first broad maximum, which appears to be due to real intermolecular interference effects, appears at about $r \sim 4$ A.,

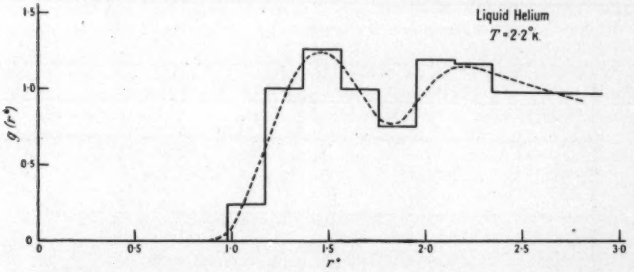


Figure 19. Experimental curves for $g(r^*)$ as function of $r^* = r/\sigma$ for liquid helium ($\sigma = 2{\cdot}56$ A.).

with $g(r) \sim 1{\cdot}6$ the following minimum occurring at about 5·5 A., with $g(r) \sim 0{\cdot}85$. Taking for σ of nitrogen the value 3·75 A. of § 8 (iii) these values occur at 1·05 and 1·5 respectively. The results for the three gases are compared in the following table:

	Argon		Nitrogen	Helium
	T_m	T_{cr}	T_m	T_λ
r^*_{max}	1·1	~1·2	1·05	1·5
$g(r^*_{max})$	3·5	1·4	1·6	1·2
r^*_{min}	1·35	1·2	1·5	1·9
$g(r^*_{min})$	0·5	0·9	0·85	0·8

In general the effect of temperature: making the maxima and minima less intense and shifting the extrema to larger values of r, is in agreement with the behaviour of the molecular pair distribution function, as calculated for gases. The shift of the maxima to larger values of r^* is also in poor agreement with the expected changes caused by quantum theories.

SYMBOLS AND NOTATION

F = free energy.
U = energy.
S = entropy.
T = temperature.
V = volume.
H = Boltzmanns' function.
k = Boltzmann's constant.
p = pressure.
n = number density of molecules.
z = fugacity.
μ = thermodynamic potential.
$\beta = 1/kT$.

R = gas constant.
N = Avogadro's number.
Q_N = configurational partition function.
Z_N = total partition function.
$B(T)$ = second virial coefficient.
$C(T)$ = third virial coefficient.

$\lambda = h/\sqrt{(2\pi mkT)}$.
$\phi(r_{ik})$ = molecular potential field.
ϵ, σ = constants of field
$\phi(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$.
m = mass of molecules.
h = constant of Boltzmann.
$\Lambda = h/\sqrt{(m\epsilon)}$.
$\Lambda^* = h/\sigma\sqrt{(m\epsilon)}$.

$\mathbf{r}_\iota$ = position vector of molecule ι.
$\mathbf{p}_\iota$ = momentum of molecule ι.
$\mathbf{c}_\iota$ = velocity of molecule ι.
$\mathbf{F}_\iota$ = force acting on molecule ι.
$\mathbf{X}_\iota$ = external force acting on molecule ι.
$\boldsymbol{\sigma}_\iota$ = wave vector of molecule ι.

$\mathbf{r}^N \equiv (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$.
$\mathbf{p}^N \equiv (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N)$.
$\mathbf{r}^h \equiv (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_h)$.
$d\mathbf{r}^h = d\mathbf{r}_1, d\mathbf{r}_2, \ldots d\mathbf{r}_h$.
$d\mathbf{r}^{N-h} = d\mathbf{r}_{h+1} \ldots d\mathbf{r}_N$.

$H(\mathbf{r}^N, \mathbf{p}^N)$ = Hamilton function.
$G(\mathbf{r}^N, \mathbf{p}^N)$ = physical quantity.
$K(\mathbf{p}^N)$ = kinetic energy.
$\Phi(\mathbf{r}^N)$ = potential energy.
$\Xi(\mathbf{r}^N)$ = virial of forces.
$\overline{F}_\iota(\mathbf{r}^h)$ = average force acting on ι in group $\{h\}$.
$\Psi(\mathbf{r}^h)$ = potential of average force.
$W(\mathbf{r}^N)$ = Slater sum.

E_σ = eigenvalue.
G_σ = expect. value.
K_σ = ,, ,,
Φ_σ = ,, ,,
Ξ_σ = ,, ,,

$P_N(\mathbf{r}^N, \mathbf{p}^N)$ = probability density.
$\sigma_N(\mathbf{r}^N, \mathbf{p}^N)$ = surface probability density.
$P_N(\mathbf{r}^N)$ = configurational probability density.
$f_N(\mathbf{r}^N, \mathbf{p}^N)$ = generic probability density.
$n_N(\mathbf{r}^N)$ = generic configurational probability density.
$P_h(\mathbf{r}^h, \mathbf{p}^h)$, $P_h(\mathbf{r}^h)$ = probability density for molecules $1 \ldots h$.
$f_h(\mathbf{r}^h, \mathbf{p}^h)$, $n_h(\mathbf{r}^h)$ = generic probability density for molecules $1 \ldots h$.
$\bar{P}_N$, $\bar{f}_h$ = coarse grained probability densities.
$\tilde{f}_N$, $\tilde{f}_h$ = time smoothed probability densities.
$\overline{G}$ = ensemble average of $G(\mathbf{r}^N, \mathbf{p}^N)$.
$\widetilde{G}$ = time average of $G(\mathbf{r}^N, \mathbf{p}^N)$.

$\psi_\sigma(\mathbf{r}^N)$ = energy eigenfunctions.
$\phi_\varrho(\mathbf{r}^N)$, $\chi_\sigma(\mathbf{r}^N)$ = arbitrary sets of orthogonal and normalized functions.
$\mathscr{H}$, $\mathscr{G}$, $\mathscr{K}$, $\mathscr{p}_\iota$ = operators corresponding to H, G, K and $\mathbf{p}_\iota$.
$\mathscr{P}$ = statistical operator.

$\mathscr{P}_{\varrho\varrho'}, \mathscr{P}_{\sigma\sigma'}, \mathscr{P}_N(\mathbf{r}^N, \mathbf{r}^{N'})$ = probability density matrix.
$n_h(\mathbf{r}^h, \mathbf{r}^{h'})$ = probability density matrix for h molecules.

$\iota, \kappa, \lambda, \mu, \nu$ = indices of molecules.
k, l, n = summation numbers.
σ, ρ, τ = quantum numbers.
$\{h\}$ = set of molecules 1, 2, ..., h.

$\mathbf{r} \equiv (x, y, z)$.
$d\mathbf{r} = dx\, dy\, dz$ = differential element of space.
$\frac{\partial}{\delta \mathbf{r}} \equiv \left(\frac{\partial}{\delta x}, \frac{\partial}{\delta y}, \frac{\partial}{\delta z}\right)$.
$\delta\mathbf{r} = \delta x \delta y \delta z$ = finite element of space.

REFERENCES

BERNAL, J. D., 1937, *Trans. Faraday Soc.*, **33**, 27.
BETH, E., and UHLENBECK, G. E., 1937, *Physica*, **4**, 915.
BIRKHOFF, G. D., 1922, *Acta Math.*, **43**, 113 ; 1931, *Proc. Nat. Acad. Sci. Wash.*, **17**, 650 ; 1932, *Ibid.*, **18**, 279.
BIRKHOFF, G. D., and SMITH, J., 1928, *J. Math.*, **7**, 365.
BLOCH, F., 1932, *Z. Phys.*, **74**, 295.
DE BOER, J., 1949 a, b, *Physica*, **15**, in press.
DE BOER, J., 1937, *Communication at the Van der Waalscongress, Amsterdam* ; 1940, *Contribution to the Theory of Compressed Gases* (Thesis, Amsterdam) ; 1948, *Physica*, **14**, 139.
DE BOER, J., and BLAISSE, B. S., 1948, *Physica*, **14**, 149.
DE BOER, J., and LUNBECK, R. J., 1948, *Physica*, **14**, 520 ; 1948, *Ibid.*, **14**, 510.
DE BOER, J., and MICHELS, A., 1938, *Physica*, **5**, 945 ; 1939, *Ibid.*, **6**, 97 ; 1939 a, *Ibid.*, **6**, 409.
BOGOLUBOV, N., 1946, *J. Phys. U.S.S.R.*, **10**, 265.
BOLTZMANN, L., 1871, *Wien. Ber.*, **63**, 679; 1882, *Phil. Mag.*, **14**, 229; 1896/98, *Vorlesungen über Gastheorie*,
BORN, M., 1933, *Optik* (Springer), § 81.
BORN, M., and FUCHS, K., 1938, *Proc. Roy. Soc.* A, **166**, 391.
BORN, M., and GREEN, H. S., 1946, *Proc. Roy. Soc.* A, **188**, 10 ; 1947 a, *Ibid.*, **190**, 455; 1947 b, *Ibid.*, **191**, 168 ; 1947 c, *Ibid.*, **192**, 166; 1948, *Ibid.*, **197**, 244.
BUCKINGHAM, R. A., HAMILTON, J., and MASSEY, H. S. W., 1941, *Proc. Roy. Soc.* A, **179**, 103.
BYK, A., 1921, *Ann. Phys., Lpz.*, **66**, 167 ; 1922, *Ibid.*, **69**, 161.
CARATHEODORY, C., 1919, *Berl. Ber.*, 580.
CHAPMAN, S., and COWLING, T. G., 1939, *The Mathematical Theory of Non-Uniform Gases* (Cambridge : University Press).
CORNER, J., and LENNARD-JONES, J. E., 1941, *Proc. Roy. Soc.* A, **170**, 40.
DARWIN, C. G. and FOWLER, R. H., 1922, *Phil. Mag.*, **44**, 450.
DEBYE, P., 1915, *Ann.Phys., Lpz.*, **46**, 809.
DEBYE, P., and MENKE, H., 1931, *Ergebnisse der Techn. Röntgenk.* II. ; *Z. Phys.*, **31**, 797.
DIRAC, P. A. M., 1929, *Proc. Camb. Phil. Soc.*, **25**, 62 ; 1930, *Ibid.*, **26**, **27**, 240, 376 ; 1935, *Quantum Mechanics* (Oxford : University Press).
EHRENFEST, P.,1906, *Wien. Ber.*, **115**, 89.
EHRENFEST, P., and EHRENFEST, T., 1911, *Encyclopaedie der Math. Wissenschaften*, IV.
EHRENFEST, P., and OPPENHEIMER, J. R., 1931, *Phys. Rev.*, **37**, 333.
EINSTEIN, A., 1910, *Ann. Phys., Lpz.*, **33**, 1275 ; 1924, *Berl. Ber.*, 261 ; 1925, *Ibid.*, 3.
EISENSCHITZ, R., 1949, *Proc. Phys. Soc.* A., **62**, 41.
EISENSTEIN, A., and GINGRICH, N. S., 1940, *Phys. Rev.*, **58**, 307 ; 1942, *Ibid.*, **62**, 261.
FOWLER, R. H., 1936, *Statistical Mechanics*, (Cambridge : University Press); 1938, *Proc. Camb. Phil. Soc.*, **34**, 302.
FOWLER, R. H., and GUGGENHEIM, E. A., 1940, *Statistical Thermodynamics* (Cambridge : University Press).

FÜRTH, R., 1941, *Proc. Camb. Phil. Soc.*, **37**, 252.
GIBBS, J. W., 1902, *Elementary Principles of Statistical Mechanica*, Collected Works (Vol. II).
GINGRICH, N. S., 1943, *Rev. Mod. Phys.*, **15**, 90.
GREEN, H. S., 1947, *Proc. Roy. Soc.* A, **189**, 103 ; 1948, Private communication.
GROPPER, L., 1937, *Phys. Rev.*, **50**, **51**, 963, 1108 ; 1939, *Ibid.*, **55**, 1095.
GUGGENHEIM, E. A., 1939, *J. Chem. Phys.*, **7**, 103 ; 1945, *Ibid.*, **13**, 253.
HIRSCHFELDER, J. O., EWELL, R. B., and ROEBUCK, J. R., 1938, *J. Chem. Phys.*, **6**, 205.
HIRSCHFELDER, J. O., and ROSEVAERE, W. E., 1939, *J. Phys. Chem.*, **43**, 15.
HIRSCHFELDER, J. O., STEVENSON, D., and EYRING, H., 1937, *J. Chem. Phys.*, **5**, 896.
HUSIMI, KODI, 1940, *Proc. Phys.-Math. Soc. Japan*, **22**, 254.
JEANS, J. H., 1925, *Kinetic Theory of Gases* (Cambridge : University Press).
KAHN, B., 1938, *Dissertation*, Utrecht.
KAHN, B., and UHLENBECK, G. E., 1938, *Physica*, **5**, 399.
KEESOM, W. H., and KRAAK, H. H., 1935, *Physica*, **2**, 37.
KEESOM, W. H., and DE SMEDT, J., 1922, *Proc. K. Acad. Amsterdam*, **25**, 118 ; 1923 *Ibid.*, **26**, 112.
KEESOM, W. H., and TACONIS, K. W., 1937, *Physica*, **4**, 28 ; 1937, *Ibid.*, **4**, 256 ; 1938, *Ibid.*, **5**, 270.
KEESOM, W. H., and WALSTRA, W. K., 1940, *Physica*, **7**, 985.
KELVIN, Lord, 1891, *Collected Works*, **4**, 484–512.
KIRKWOOD, J. G., 1933, *Phys. Rev.*, **44**, 31 ; 1935, *J. Chem. Phys.*, **3**, 300 ; 1936, *Ibid.*, **4**, 592 ; 1939, *Ibid.*, **7**, 919; 1946, *Ibid.*, **14**, 180 ; 1947, *Ibid.*, **15**, 72.
KIRKWOOD, J. G., and BOGGS, E. M., 1941, *J. Chem. Phys.*, **10**, 394.
KISTENMAKER, J., and KEESOM, W. H., 1946, *Physica*, **12**, 227.
KOOPMAN, B. O., 1931, *Proc. Nat. Acad. Sci. Wash.*, **17**, 315.
LANGEVIN, P., 1908 *C.R. Acad. Sci. Paris*, **146**, 530.
LENNARD-JONES, J. E., 1924, *Proc. Roy. Soc.* A, **106**, 463 ; 1931, *Proc. Phys. Soc.*, **43**, 461 ; 1937, *Physica*, **4**, 941.
LENNARD-JONES, J. E., and DEVONSHIRE, A. F., 1937, *Proc. Roy. Soc.* A, **163**, 53; 1938, *Ibid.*, **165**, 1.
LONDON, F., 1938, *Phys. Rev.*, **54**, 947 ; 1938, *Nature, Lond.*, **141**, 643.
LORENTZ, H. A., 1907, *Abh. Theor. Phys.*, XI, § 9.
MASSEY, H. S. W., and BUCKINGHAM, R. A., 1938, *Proc. Roy. Soc.* A, **168**, 378 ; 1939, *Ibid.*, **169**, 205.
MAXWELL, J. C., 1879, *Trans. Camb. Phil. Soc.*, **12**, 547.
MAYER, J. E., 1937, *J. Chem. Phys.*, **5**, 67 ; 1942, *J. Chem. Phys.*, **10**, 629 ; 1947, *Ibid.*, **15**, 187.
MAYER, J. E., and BAND, W., 1947, *J. Chem. Phys.*, **15**, 141.
MAYER, J. E., and HARRISON, S. F., 1938, *J. Chem. Phys.*, **6**, 87, 101.
MAYER, J. E., and MAYER, M. G., 1940, *Statistical Mechanics* (New York : Wiley and Sons).
MAYER, J. E., and MONTROLL, E. W., 1941, *J. Chem. Phys.*, **9**, 2.
MICHELS, A., BIJL, A., and MICHELS, C., 1937 c, *Proc. Roy. Soc.* A, **160**, 376.
MICHELS, A., BLAISSE, B., and MICHELS, C., 1937 b, *Proc. Roy. Soc.* A, **160**, 358.
MICHELS, A., and GELDERMANS, M., 1942, *Physica*, **9**, 967.
MICHELS, A., and GOUDEKET, M., 1941 a, *Physica*, **8**, 347 ; 1941 b, *Ibid.*, **8**, 353 ; 1941 c, *Ibid.*, **8**, 387.
MICHELS, A., and MICHELS, C., 1937 a, *Proc. Roy. Soc.* A, **160**, 348.
MICHELS, A., WOUTERS, H., and DE BOER, J., 1936 a, *Physica*, **3**, 585 ; 1936 b, *Ibid.*, **3**, 597.
MONTROLL, E. W., and MAYER, J. E., 1941, *J. Chem. Phys.*, **9**, 626.
VAN NEUMANN, J., 1927, *Gött. Nachr.*, 245 ; 1932 a, *Proc. Nat. Acad. Sci. Wash.*, **18**, 70 ; 1932 b, *Math. Grundlagen, der Quatenmechanik* (Berlin).
ONSAGER, L., 1933, *Chem. Rev.*, **13**, 73.
ORNSTEIN, L. S., 1908, *Thesis*, Utrecht.
ORNSTEIN, L. S., and ZERNIKE, F., 1914, *Proc. K. Acad. Amsterdam*, **17**, 793; 1926, *Phys. Z.*, **27**, 761.
PAULI, W., 1927, *Z. Phys.*, **41**, 81.
PITZER, K. S., 1939, *J. Chem. Phys.*, **7**, 583.
PLANCHEREL, M., and ROSENTHAL, A., 1913, *Ann. Phys., Lpz.*, **42**, 796, 1061.

PAGE G

POINCARE, H., 1894, *Rev. Gen. des Sciences*, 516.
PRINS, J. A., 1937, *Trans. Faraday Soc.*, **33**, 110.
RUSHBROOKE, C. S., and COULSON, C. A., 1939, *Phys. Rev.*, **56**, 1216.
SCHÄFER, KL., 1937, *Z. Phys. Chem.* B, **36**, 85 ; *Ibid.*, **38**, 187.
SHARRAH, P. C., and GINGRICH, N. S., *J. Chem. Phys.*, **10**, 504.
SLATER, J. C., 1931, *Phys. Rev.*, **38**, 237 ; 1933, *J. Chem. Phys.*, **1**, 687.
SMOLUCHOWSKI, M., 1908, *Ann. Phys., Lpz.*, **25**, 205.
TOLMAN, R., 1938, *Statistical Mechanics* (Oxford: University Press); 1940, *Phys. Rev.*, **57**, 1160.
UHLENBECK, G. E., 1927, *Dissertation*, Leiden ; 1935, *J. Math. Phys.*, **14**, 10.
UHLENBECK, G. E., and BETH, E., 1936, *Physica*, **3**, 729.
UHLENBECK, G. E., and GROPPER, L., 1932, *Phys. Rev.*, **41**, 79.
URSELL, H. D., 1927, *Proc. Camb. Phil. Soc.*, **23**, 685.
WALL, C. N., 1940, *Phys. Rev.*, **50**, 307.
WIGNER, E., 1932, *Phys. Rev.*, **40**, 749.
YVON, J., 1935, *Actualitiés Scientifiques*, No. 203 ; 1937 a, *Ibid.*, 542 ; 1937 b, *Ibid.*, 543.
ZERNIKE, F., and PRINS, J. A., 1927, *Z. Phys.*, **41**, 184.

REPORTS ON PROGRESS IN PHYSICS

Vol. 1 (1934). Pp. 371. Price 30*s*. (*Out of print.*)
Articles by: G. Temple, G. C. McVittie, A. Ferguson, G. F. P. Trubridge, E. G. Richardson, R. W. B. Pearse, H. Dingle, E. Gwynne Jones, C. V. Jackson, A. Hunter, T. Smith, J. S. Anderson, J. Guild, H. F. Buckley, J. H. Awbery, J. C. McLennan, G. E. Bell, I. Backhurst, W. Binks, G. Shearer, W. A. Wood, J. Thewlis, E. N. da C. Andrade, L. Hartshorn.

Vol. 2 (1935). Pp. 371. Price 30*s*. (*Out of print.*)
Articles by: A. Ferguson, G. Temple, R. Peierls, G. B. B. M. Sutherland, W. G. Penney, R. Schlapp, N. Feather, E. Lancaster-Jones, E. V. Appleton, E. G. Richardson, J. H. Awbery, L. Hartshorn, G. I. Finch, J. L. Miller, J. E. L. Robinson, H. J. H. Starks, G. E. Bell, I. Backhurst, W. Binks, W. E. Perry, G. Shearer, W. A. Wood, J. Thewlis, C. V. Jackson, R. W. B. Pearse, A. Hunter, Lotte Kellner, E. Gwynne-Jones, T. Smith, J. Guild.

Vol. 3 (1936). Pp. 394. Price 30*s*. (*Out of print.*)
Articles by: R. C. Brown, H. L. Cox, D. G. Sopwith, G. Barr, A. Fage, S. Chapman, W. C. Price, N. Feather, A. Nunn May, E. G. Richardson, G. W. C. Kaye, R. W. Powell, L. F. Bates, L. Hartshorn, D. A. Oliver, C. E. Webb, P. Vigoureux, C. E. Wynn-Williams, A. H. Wilson, G. A. Boutry, L. C. Martin, B. K. Johnson, F. Simeon, W. Swaine, W. D. Wright, W. S. Stiles, J. T. Randall, G. E. Bell, I. Backhurst, W. Binks, W. A. Wood, W. E. T. Perry, W. Jevons, M. F. Crawford, A. Hunter, R. W. B. Pearse, W. H. J. Childs.

Vol. 4 (1937). Pp. 396. Price 30*s*. (*Now available.*)
Articles by: H. Spencer Jones, E. K. Rideal, R. C. Brown, E. G. Richardson, R. W. Powell, J. H. Awbery, E. Griffiths, H. T. Flint, C. H. Collie, P. B. Moon, H. R. Robinson, L. Hartshorn, T. I. Jones, W. H. Ward, R. W. Gurney, F. Simeon, W. Jevons, W. R. S. Garton, A. Hunter, R. W. B. Pearse, E. W. Foster, Lotte Kellner, J. M. Robertson, K. Lonsdale.

Vol. 5 (1938). Pp. 445. Price 30*s*. (*Out of print.*)
Articles by: R. C. Brown, H. L. Cox, C. F. Goodeve, J. W. McBain, G. F. Mills, N. F. Mott, R. W. Gurney, H. W. Melville, A. C. D. Crommelin, D. Brunt, E. G. Richardson, P. M. Tookey Kerridge, R. W. Powell, L. C. Martin, W. M. Hampton, W. N. Wheat, E. W. H. Selwyn, W. D. Wright, W. Jevons, A. G. Shenstone, M. F. Crawford, J. Stuart Foster, A. Hunter, R. W. B. Pearse, H. W. B. Skinner, W. V. Mayneord, L. Hartshorn, N. F. Astbury, F. C. Champion, W. Heitler, A. Nunn May, H. T. Flint, A. W. Barton.

Vol. 6 (1939). Pp. 434. Price 30*s*. (*Out of print.*)
Articles by: H. A. Bethe, R. E. Marshak, H. J. Walke, H. C. Urey, G. W. C. Kaye, G. E. Bell, W. Binks, W. E. Perry, W. B. Mann, W. Jevons, A. Hunter, H. E. White, F. H. Crawford, J. T. Randall, M. H. F. Wilkins, N. F. Mott, W. G. Penney, E. G. Richardson, A. Fage, H. Jones, R. W. Powell, H. Pelzer, L. C. Jackson, J. M. Robertson, W. H. Taylor, H. Lipson, E. T. Goodwin, L. Hartshorn, R. S. Rivlin, N. F. Astbury, H. Fröhlich, D. Owen.

Vol. 7 (1940). Pp. 362. Price 30*s*. (*Out of print.*)
Articles by: E. G. Richardson, W. D. Wright, H. H. Nielsen, N. Feather, R. Peierls, O. Klemperer, H. W. Lee, A. Hunter, A. D. Thackeray, R. C. Brown, J. A. Beattie, W. H. Stockmayer, V. J. Francis, H. G. Jenkins, J. K. Roberts, F. Eirich, A. Ferguson.

Vol. 8 (1941). Pp. 372. Price 30*s*. (*Out of print.*)
Articles by: M. Born, M. Blackman, J. W. Beams, A. G. Gaydon, R. Schnurmann, R. T. Birge, V. K. Zworykin, R. E. Shelby, A. Hunter, Lotte Kellner, G. R. Harrison, R. S. Mulliken, C. A. Rieke, E. P. Wigner, E. Feenberg, J. M. Meek.

Vol. 9 (1942–43). Pp. 353. Price 30*s*. (*Out of print.*)
Articles by: S. Chapman, A. Hunter, W. C. Price, G. B. B. M. Sutherland, G. S. Callender, T. G. Cowling, R. W. B. Pearse, J. Sayers, H. S. W. Massey, D. R. Bates, C. H. Bamford, L. R. G. Treloar, J. D. Craggs, A. E. Alexander, R. O. Jenkins, J. McG. Bruckshaw, E. G. Richardson, K. Lonsdale, M. Born, F. I. G. Rawlins.

Vol. 10 (1944–45). Pp. 442. Price 30*s*. (*Now available.*)
Articles by: W. F. G. Swann, H. Jeffreys, H. P. Rooksby, E. G. Richardson, P. Koets, A. B. D. Cassie, J. G. Martindale, L. Marton, H. J. Bhabha, A. F. C. Pollard, J. M. Meek, F. R. Perry, K. Mendelssohn, J. Cabannes, J. D. Bernal.

Vol. 11 (1946–47). Pp. 461. Price 42*s*. (*Now available.*)
Articles by: R. J. Van de Graaff, J. G. Trump and W. W. Buechner; N. Feather; B. Porntecorvo; Edmund C. Stoner; D. R. Hartree; G. B. B. M. Sutherland and E. Lee; B. Bleaney; Charles Kittel; W. F. Berg; H. Friedenstein, S. L. Martin and G. L. Munday; H. Moore; H. L. Penman; A. C. B. Lovell, J. P. M. Prentice, J. G. Porter, R. W. B. Pearse and N. Herlofson.

Published by
The Physical Society, 1 Lowther Gardens, Prince Consort Road, S.W.7

Orders with remittance should be sent to the above address